

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 9, 2003, 16:48:09 ; Search time 53 Seconds

(without alignments)  
4306.581 Million cell updates/sec

Title: US-10-006-091-1

Perfect score: 7691

Sequence: 1 ATRYYLGAVELSWDYMQSD.....WVHQIALRMEVLGCEAQQDLY 1438

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_19Jun03.\*  
1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7691	100.0	1438	21 AAB01262	B-domain deleted f
2	7691	100.0	1457	20 AAY21675	Beta-domain delete
3	7684	99.9	1457	19 AAW46246	Human factor VIII
4	7684	99.9	1457	19 AAW44372	Human Factor VIII
5	7674	99.8	1471	18 AAW23414	Human B-domain del
6	7674	99.8	1471	22 AAB67959	Amino acid sequenc
7	7659.5	99.6	1440	12 AAR12971	Factor VIII:SQ. U
8	7626	99.2	1459	22 AAE10832	Human factor VIII
9	7616	99.0	1459	22 AAE10833	Human factor VIII

10	7613	99.0	1459	22 AAE10827	Human factor VIII
11	7611.5	99.0	1516	9 AAP80285	Modified factor VI
12	7592	98.7	1424	22 AAB48842	Mutant mature huma
13	7592	98.7	1424	23 AAO18622	Human mature B-dom
14	7588	98.7	1447	23 ABG92541	SArg B-domain-dele
15	7581	98.6	1425	9 AAP80267	Modified factor VI
16	7578	98.5	1424	9 AAP80268	Modified factor VI
17	7578	98.5	1445	23 ABG92540	LE B-domain-delete
18	7574	98.5	1424	10 AAP91169	Sequence of 740 Ar
19	7434	96.7	1661	18 AAW18670	Factor VIII-dB695-
20	7353.5	95.6	1383	18 AAW33229	Procoagulant-activ
21	7345.5	95.5	1383	18 AAW33229	Procoagulant-activ
22	7342.5	95.5	1383	18 AAW33228	Procoagulant-activ
23	7234	94.1	2332	14 AAW43257	Human factor VIII.
24	7234	94.1	2332	19 AAW53483	Human factor VIII.
25	7234	94.1	2332	20 AAY31594	Human factor VIII
26	7234	94.1	2332	22 AAE10826	Human mature wild-
27	7234	94.1	2332	22 AAE11200	Human factor VIII
28	7234	94.1	2332	22 AAE50465	Human factor VIII.
29	7234	94.1	2332	23 AAU79869	Human factor VIII.
30	7234	94.1	2351	20 AAY21676	Factor VIII protei
31	7231.5	94.0	2342	18 AAW11422	Active Factor VIII
32	7231	94.0	2351	18 AAW10591	Factor VIII:C (Phe
33	7231	94.0	2351	18 AAW13496	Factor VIII:C (Arg
34	7230.5	94.0	2344	18 AAW11432	Active Factor VIII
35	7230.5	94.0	2344	18 AAW11410	Active Factor VIII
36	7230	94.0	2332	23 AAU79870	Human factor VIII
37	7230	94.0	2332	23 AAU79872	Human factor VIII
38	7230	94.0	2345	18 AAW11415	Active Factor VIII
39	7230	94.0	2345	18 AAW11403	Active Factor VIII
40	7229.5	94.0	2346	18 AAW11421	Active Factor VIII
41	7229.5	94.0	2346	18 AAW11431	Active Factor VIII
42	7229.5	94.0	2346	18 AAW11434	Active Factor VIII
43	7229	94.0	2347	18 AAW11411	Active Factor VIII
44	7229	94.0	2347	18 AAW11402	Active Factor VIII
45	7229	94.0	2351	18 AAW10592	Factor VIII:C (Tyr

#### ALIGNMENTS

RESULT 1  
AAB01262 ID AAB01262 standard; protein; 1438 AA.

AC AAB01262;

DT 25-SEP-2000 (first entry)

DE B-domain deleted factor VIII sequence.

Factor VIII; procoagulant; adenovirus; adeno-associated strain;  
gene therapy; human Burkitt's lymphoma; HKB; therapy;  
therapeutic protein; vector; Epstein-Barr virus; human.

OS Homo sapiens.

PN WO200034505-A1.

PD 15-SEP-2000.

PF 08-DEC-1999; 99WO-US29169.

PR 10-DEC-1998; 98US-0209916.

PA (FARB ) BAYER CORP.

PI Cho M, Chan SY, Kelsey W, Yee H;

DR WPI; 2000-431311/37.

PT Producing cells expressing a protein having factor VIII procoagulant activity especially, human factor VIII in an industrial scale, involves

PT expressing a vector comprising a sequence coding for factor VIII in  
PT human cells  
XX  
PS Claim 7; Fig 1; 27pp; English.  
XX  
Producing cells expressing a protein having factor VIII procoagulant  
CC activity, comprises contacting the cells with a vector comprising a  
CC selectable marker and a sequence coding for the protein having factor  
CC VIII procoagulant activity operably linked to a promoter. The cells  
CC are then selected and individual clones expressing high levels of the  
CC protein are isolated from the selected cells. The cells produced by  
CC the method are not only useful for producing protein having factor  
CC VIII procoagulant activity but also for producing adenovirus and  
CC adeno-associated virus strains for gene therapy. The advantage of  
CC having cells producing protein with factor VIII procoagulant activity  
CC is that factor VIII protein can be produced on an industrial scale  
CC in the range of 2-4 pg/cell/day. Human Burkitt's lymphoma (HKB)  
CC cells provide a protein-free production system to produce not only  
CC B-domain deleted factor VIII but also other therapeutic proteins. The  
CC vector used in the method preferably comprises B-domain deleted  
CC factor VIII (BDD-FVIII), a transcriptional unit for BDD-FVIII and a  
CC selectable marker, dihydrofolate reductase (dhfr). In addition, a  
CC terminal repeat sequence from Epstein-Barr virus is inserted into the  
CC vector to increase integration efficiency.  
XX  
SQ Sequence 1438 AA;  
Query Match 100.0%; Score 7691; DB 21; Length 1438;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATRRYLGAVELSWDMQSDLGELPVDARFPVRPKSPFNTSVYKKTFLVEFVTHLEN 60  
DB 1 ATRRYLGAVELSWDMQSDLGELPVDARFPVRPKSPFNTSVYKKTFLVEFVTHLEN 60  
QY 61 IAKPRPPMGLLGPTIQAQVYDTVTITLKNASHPVSLHAGVSVKASEGAEYDDQTSQ 120  
DB 61 IAKPRPPMGLLGPTIQAQVYDTVTITLKNASHPVSLHAGVSVKASEGAEYDDQTSQ 120  
QY 121 REKEDKVPFGSGSHYVMQVLKENGPMASDPLCLTYSLYSHVDLVKDLNSGLIGALLVCR 180  
DB 121 REKEDKVPFGSGSHYVMQVLKENGPMASDPLCLTYSLYSHVDLVKDLNSGLIGALLVCR 180  
QY 181 EGS LAKEKTQTLHKFTLLFAVDEGKSWHSETKNSLMQDRDAASARAWPKMHTVANGYNR 240  
DB 181 EGS LAKEKTQTLHKFTLLFAVDEGKSWHSETKNSLMQDRDAASARAWPKMHTVANGYNR 240  
QY 241 SLPLGIGCHRSVYWHVIGMGTTPEVHSIFLEGHTFLVRNHRQASLEISPTITFLTAQTLL 300  
DB 241 SLPLGIGCHRSVYWHVIGMGTTPEVHSIFLEGHTFLVRNHRQASLEISPTITFLTAQTLL 300  
QY 301 MDLGOFLLFCHTSSHQHDMAYVKVDCSPESPQLRMKNNEAEYDDDLTDSMDVVRP 360  
DB 301 MDLGOFLLFCHTSSHQHDMAYVKVDCSPESPQLRMKNNEAEYDDDLTDSMDVVRP 360  
QY 361 DDNSPSFIQIRSAVKHKPTWVHYTAAEEEDWDYAPLVLAAPDRDSYKSQYLNNGPQRIG 420  
DB 361 DDNSPSFIQIRSAVKHKPTWVHYTAAEEEDWDYAPLVLAAPDRDSYKSQYLNNGPQRIG 420  
QY 421 RYKXKVRFMAYTDTFTKTRATOHESGILGPLLYGEVGDTLIIIPKQASRPYNIYPHGI 480  
DB 421 RYKXKVRFMAYTDTFTKTRATOHESGILGPLLYGEVGDTLIIIPKQASRPYNIYPHGI 480  
QY 481 TDVRPLYSRRLPKGVKHLKDFILPGEIPIKYKWTVTVEDGPKSDPRCLTRYYSFVNME 540  
DB 481 TDVRPLYSRRLPKGVKHLKDFILPGEIPIKYKWTVTVEDGPKSDPRCLTRYYSFVNME 540  
QY 541 RLASGLIGPLLI CYKESVQDQGNQIMSKRNVILFVSFDEKRSWYLTENIKRFLPNPAG 600  
DB 541 RLASGLIGPLLI CYKESVQDQGNQIMSKRNVILFVSFDEKRSWYLTENIKRFLPNPAG 600  
QY 601 VOLEDPEFOASNIMHSINGYVFDLSQLSVCLHEVAYWYILSIGAQDTFLSVFFSGYTFKH 660  
DB 601 VOLEDPEFOASNIMHSINGYVFDLSQLSVCLHEVAYWYILSIGAQDTFLSVFFSGYTFKH 660

DB 601 VOLEDPEFOASNIMHSINGYVFDLSQLSVCLHEVAYWYILSIGAQDTFLSVFFSGYTFKH 660  
QY 661 KMVYEDTLTLFPFSGETVFMSENPGLMWILGCHNSDFRNGMTALLKVSSCDKNTGDYYE 720  
DB 661 KMVYEDTLTLFPFSGETVFMSENPGLMWILGCHNSDFRNGMTALLKVSSCDKNTGDYYE 720  
QY 721 DSYEDISAYLLSKNNAIBPRFSQNPVVLKXHQREITRTTLQSDQEEIDYDDTISVEMKK 780  
DB 721 DSYEDISAYLLSKNNAIBPRFSQNPVVLKXHQREITRTTLQSDQEEIDYDDTISVEMKK 780  
QY 781 EDFDIYDEDENQSPRSFQKKTTHYFIAAVERLWDYGMSSPHVLNRQAQSVPOFKKVV 840  
DB 781 EDFDIYDEDENQSPRSFQKKTTHYFIAAVERLWDYGMSSPHVLNRQAQSVPOFKKVV 840  
QY 841 FOEFDTGSGFTOPLVRGELNEHLGLLGPYIRAEVDENIMVTFRNOASRPVSVSSLSIYEE 900  
DB 841 FOEFDTGSGFTOPLVRGELNEHLGLLGPYIRAEVDENIMVTFRNOASRPVSVSSLSIYEE 900  
QY 901 DQRQGAEPKRFVFPNETKTYFWKVQHMAPTKDEFCCKAWAYFSDVLEKDVHSGLIGP 960  
DB 901 DQRQGAEPKRFVFPNETKTYFWKVQHMAPTKDEFCCKAWAYFSDVLEKDVHSGLIGP 960  
QY 961 LLVCHTNTLNPAHGRQVTVQEFALFTTIDETKSWYFTENMERNCRAPCNQIMEDPTPKE 1020  
DB 961 LLVCHTNTLNPAHGRQVTVQEFALFTTIDETKSWYFTENMERNCRAPCNQIMEDPTPKE 1020  
QY 1021 NYRFAINGYIMDTLPGLVMAQDQIRWYLLSMGSENHISHPSGHVFTVRKKEEYKMA 1080  
DB 1021 NYRFAINGYIMDTLPGLVMAQDQIRWYLLSMGSENHISHPSGHVFTVRKKEEYKMA 1080  
QY 1081 LYNLYPGVFETVEMLPKSGAGIWRVECLIGEHLHAGMSTFLVYSNKCOPTPLGMAASHIRD 1140  
DB 1081 LYNLYPGVFETVEMLPKSGAGIWRVECLIGEHLHAGMSTFLVYSNKCOPTPLGMAASHIRD 1140  
QY 1141 FQITASGOYGOWAPKPLARLHYSGSINAWSTKEPSPWIKVDLLAPMIHGIKTQGAROKFS 1200  
DB 1141 FQITASGOYGOWAPKPLARLHYSGSINAWSTKEPSPWIKVDLLAPMIHGIKTQGAROKFS 1200  
QY 1201 SLYTSQFIIMYSLDGKKWQTVRGSTGTLMVFFGNVDSGSGIKHNI FNPPIARYIRLHPT 1260  
DB 1201 SLYTSQFIIMYSLDGKKWQTVRGSTGTLMVFFGNVDSGSGIKHNI FNPPIARYIRLHPT 1260  
QY 1261 HYSIRSTLRMBELMGCDLNSCSMPLGMSKASIDAQITASSYFTNNFATWSPSKARLHLQ 1320  
DB 1261 HYSIRSTLRMBELMGCDLNSCSMPLGMSKASIDAQITASSYFTNNFATWSPSKARLHLQ 1320  
QY 1321 RSNARPOVNNPKEWLQYDFQKTMKVTVGTTQGVKSLTSMYVKEFLISSQDGHQWTLF 1380  
DB 1321 RSNARPOVNNPKEWLQYDFQKTMKVTVGTTQGVKSLTSMYVKEFLISSQDGHQWTLF 1380  
QY 1381 FQNGKVKVFOGNDQSFPTVNSLDPPLLTRYLRIHPQSWVHQIALRMVILGCEAODLY 1438  
DB 1381 FQNGKVKVFOGNDQSFPTVNSLDPPLLTRYLRIHPQSWVHQIALRMVILGCEAODLY 1438  
RESULT 2  
AAV21675  
ID AAV21675 standard; Protein; 1457 AA.  
XX  
AC AAV21675;  
XX  
DT 18-AUG-1999 (first entry)  
XX  
DE Beta-domain deleted Factor VIII protein.  
XX  
KW Factor VIII protein; gene modification; gene therapy; clinical disorder;  
KW splicing pattern; RNA processing; gene regulation; beta-domain; human.  
XX Homo sapiens.  
XX  
PN WO9929848-A1.  
XX  
PD 17-JUN-1999.

XX PF 25-NOV-1998; 98WO-US25354.  
XX PR 16-JAN-1998; 98US-0071596.  
XX PR 05-DEC-1997; 97US-0067614.  
XX PA (IMMU-) IMMUNE RESPONSE CORP.  
XX PI Bidlingmaier S, Gonzales JEN, Ill CR, Yang CQ;  
XX DR WPI; 1999-385602/32.  
XX DR N-PSDB; AAX82258, AAX82259, AAX82260.  
XX PT Genes and vectors exhibiting increased expression and novel splicing  
XX PT patterns, useful for expression of, e.g. beta-domain deleted factor  
XX PT VIII  
XX PS Disclosure; Page 72-78; 123pp; English.  
XX CC The invention describes novel genes and vectors exhibiting increased  
XX CC expression and novel splicing patterns. It provides a gene encoding a  
XX CC Factor VIII protein, that comprises one or more consensus or near  
XX CC consensus splice sites which have been corrected to increase expression.  
XX CC The method DNA sequences and expression vectors can be used to increase  
XX CC the expression of a gene, especially a Factor VIII gene. Genes containing  
XX CC modified 5' and/or 3' untranslated regions have optimized expression  
XX CC levels and tissue-specific expression. The methods are used for  
XX CC identification and correction of consensus splice sites, addition of  
XX CC introns, optimization of 5' and 3' untranslated regions and increase in  
XX CC cytoplasmic RNA accumulation. Hence the DNAs are useful in gene therapy  
XX CC to treat a clinical disorder, to study RNA processing and/or gene  
XX CC regulation. The present sequence represents a beta-domain deleted Factor  
XX CC VIII protein.  
XX SQ Sequence 1457 AA;  
Query Match 100.0%; Score 7691; DB 20; Length 1457;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATRRYLGAVELSWDMQSDGLPVDARFPVRPKSPFNTSVYKTLFVEFTVHLFN 60  
DB 20 ATRRYLGAVELSWDMQSDGLPVDARFPVRPKSPFNTSVYKTLFVEFTVHLFN 79  
QY 61 IAKPRPPWMLGLPTTQAEVYDTVTITLKNMASHPVSLHAGVSYWKASGEARYDDQTSQ 120  
DB 80 IAKPRPPWMLGLPTTQAEVYDTVTITLKNMASHPVSLHAGVSYWKASGEARYDDQTSQ 139  
QY 121 REKEDDKVPFGSGHTYVQVLKENGPMASDPLCTYSYLSHVDLVKDLNSGLIGALLVCR 180  
DB 140 REKEDDKVPFGSGHTYVQVLKENGPMASDPLCTYSYLSHVDLVKDLNSGLIGALLVCR 199  
QY 181 EGS LAKEKTQTLHKFTLLFAVDEGKSWSETKNSLMQDRDASARAWPKMTVNGYVNR 240  
DB 200 EGS LAKEKTQTLHKFTLLFAVDEGKSWSETKNSLMQDRDASARAWPKMTVNGYVNR 259  
QY 241 SLPLGLIGCHRSYVHVIWGTTTPEVHSIFLEGTFLVNRHQASLEISPTIFLTAQTLL 300  
DB 260 SLPLGLIGCHRSYVHVIWGTTTPEVHSIFLEGTFLVNRHQASLEISPTIFLTAQTLL 319  
QY 301 MDLGQFLPCHISSHQDGMAYKVVDSCPEEPQLRMKNNEAEYDDDLTDSMDVVRP 360  
DB 320 MDLGQFLPCHISSHQDGMAYKVVDSCPEEPQLRMKNNEAEYDDDLTDSMDVVRP 379  
QY 361 DDNSPSFTQIRSVAKHPTWVHYTAAREEDWDYAPLVLAPDDRYSKYQLNNGPQRIG 420  
DB 380 DDNSPSFTQIRSVAKHPTWVHYTAAREEDWDYAPLVLAPDDRYSKYQLNNGPQRIG 439  
QY 421 RYKVKRFMAYTDETKTREAIQHESGILGPLLYGVBGDTLLIIFKNQASRPNIYPHGI 480  
DB 440 RYKVKRFMAYTDETKTREAIQHESGILGPLLYGVBGDTLLIIFKNQASRPNIYPHGI 499  
QY 481 TDVRPLYSRRLPGVKHKLKDFPILPGEIFKYKWTVTVEDEGPTKSDPRCLTRYSSFVNME 540

DB 500 TDVRPLYSRRLPGVKHKLKDFPILPGEIFKYKWTVTVEDEGPTKSDPRCLTRYSSFVNME 559  
QY 541 RDLASGLIGPLLIICYKESVDQGNQIMSKRNVILFSVDEENRSWYLTENIQFLPNPAG 600  
DB 560 RDLASGLIGPLLIICYKESVDQGNQIMSKRNVILFSVDEENRSWYLTENIQFLPNPAG 619  
QY 601 VOLEDEFOASNMHSINGYVFDLSQLSVCLHEVAYWYILSIGAQTDFLSVPFSGYTPKH 660  
DB 620 VOLEDEFOASNMHSINGYVFDLSQLSVCLHEVAYWYILSIGAQTDFLSVPFSGYTPKH 679  
QY 661 KMVYEDTLTLPFSGETFMENPCLWILGCHNSDFRNRGMTALLKVSSCDKNTGDYBE 720  
DB 680 KMVYEDTLTLPFSGETFMENPCLWILGCHNSDFRNRGMTALLKVSSCDKNTGDYBE 739  
QY 721 DSYEDISAYLLSKNNAIEPRSPQNPVLKRQREITRTTLOSDEIDEYDDTISVEMKK 780  
DB 740 DSYEDISAYLLSKNNAIEPRSPQNPVLKRQREITRTTLOSDEIDEYDDTISVEMKK 799  
QY 781 EDFDIYDEENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNRNRAQSGVPQPKVY 840  
DB 800 EDFDIYDEENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNRNRAQSGVPQPKVY 859  
QY 841 FQEFDTGSGTQPLYRGELNEHLGLLGPYIRAEVEDNIMVTFRNOASRPYSFSSLSIYBE 900  
DB 860 FQEFDTGSGTQPLYRGELNEHLGLLGPYIRAEVEDNIMVTFRNOASRPYSFSSLSIYBE 919  
QY 901 DORQAGPEPRKFNPKNETKYFWKQVHMAPTKDFDKAWAYFSDVDLEKDVHSLGIGP 960  
DB 920 DORQAGPEPRKFNPKNETKYFWKQVHMAPTKDFDKAWAYFSDVDLEKDVHSLGIGP 979  
QY 961 LLVCHTNTLPAHGQVTVQEPALFPTTIFDETKSMYFTENMERNCRAPCNIQMEDPTFKE 1020  
DB 980 LLVCHTNTLPAHGQVTVQEPALFPTTIFDETKSMYFTENMERNCRAPCNIQMEDPTFKE 1039  
QY 1021 NYRPHAINGYIMDTLPLVMAQDQIRWYLLSGNSNENIHSIFSGHVTVRKKEBYKMA 1080  
DB 1040 NYRPHAINGYIMDTLPLVMAQDQIRWYLLSGNSNENIHSIFSGHVTVRKKEBYKMA 1099  
QY 1081 LYNLPVGPVETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPLGWASGHIRD 1140  
DB 1100 LYNLPVGPVETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPLGWASGHIRD 1159  
QY 1141 FOITASGOYGOWAPKLARLHYSGSINANSTKEPPFWIKVDLLAPMLIHGIKTOGARQKFS 1200  
DB 1160 FOITASGOYGOWAPKLARLHYSGSINANSTKEPPFWIKVDLLAPMLIHGIKTOGARQKFS 1219  
QY 1201 SLYISQFIIMYSLDGKKQTYRGNSTGTLMVFGNVDSGGIKHNIIPNPPIIARYIRLHPT 1260  
DB 1220 SLYISQFIIMYSLDGKKQTYRGNSTGTLMVFGNVDSGGIKHNIIPNPPIIARYIRLHPT 1279  
QY 1261 HYSIRSTLRMELMGCDLNSCMPLGMESKAISDAQITASSYFTNMPATWSPSKARLHLOG 1320  
DB 1280 HYSIRSTLRMELMGCDLNSCMPLGMESKAISDAQITASSYFTNMPATWSPSKARLHLOG 1339  
QY 1321 RSNAMRPQVNNPKELQVDFQTKMVTGVTTOGVKSLTSMYVKEPFLISSQDGHQWTLF 1380  
DB 1340 RSNAMRPQVNNPKELQVDFQTKMVTGVTTOGVKSLTSMYVKEPFLISSQDGHQWTLF 1399  
QY 1381 FQNGKVKVFGNQDQSTFPVNSLDPLLTRYRIHPQSWVHQAIALRMEVLGCEADQLY 1438  
DB 1400 FQNGKVKVFGNQDQSTFPVNSLDPLLTRYRIHPQSWVHQAIALRMEVLGCEADQLY 1457  
RESULT 3  
AAX46246  
ID AAX46246 standard; Protein; 1457 AA.  
XX  
AC AAX46246;  
XX  
DT 25-MAR-2003 (updated)  
DT 06-AUG-1998 (first entry)  
XX

Human factor VIII beta-domain deleted SQN deletion protein sequence.

DE XX Replication defective; recombinant retrovirus; RRV; therapeutic protein; 139  
 KW haemophilia; thrombosis; hypercoagulable disorder; liver disease; human; 180  
 KW hepatitis; thalassemia; phenylketonuria; Lesch-Nyhan syndrome; diabetes; 199  
 KW cystic fibrosis; Duchenne's Muscular Dystrophy; hypercholesterolemia; 240  
 KW hypopituitarism; adenine deaminase deficiency; HIV infection; anaemia; 259  
 KW Guacher's syndrome; high blood pressure; Alzheimer's disease; autoimmune; 300  
 XX inflammatory disease; factor VIII.

OS Homo sapiens.

PN WO9800541-A2.

XX 08-JAN-1998.

XX 02-JUL-1997; 97WO-US11784.

XX 04-JUN-1997; 97US-0869309.

XX 03-JUL-1996; 96US-0645601.

XX 13-AUG-1996; 96US-0696381.

XX (CHIR ) CHIRON CORP.

PI Jolly DJ, Barber JR, Chang SMW, Respass JG, Allen JR, Boder M;

PI Chong X, De La Vega D, Depolo NJ, Hsu DC, Ibanez CE;

PI Mittelstaedt DM, Prusseak CE, Greengard J, Lee R;

XX WPI; 1998-086966/08.

XX N-PSDB; AAV19581.

PT New replication defective recombinant retro-viruses - which can be  
 PT administered to provide long term systemic expression of therapeutic  
 PT protein in blood, useful in, e.g. treating hyper-coagulable  
 PT disorders

XX Example 28; Pages 213-217; 272pp; English.

PS This is the beta-domain deleted SQN deletion protein of human factor  
 CC VIII. The encoding DNA is used to construct recombinant retroviral  
 CC vectors expressing human factor VIII. The invention provides the  
 CC preparation of replication defective recombinant retrovirus (RRV)  
 CC expressing a therapeutic protein. The RRV preparation is resistant to  
 CC degradation by human complement and is capable of inducing long term  
 CC systemic expression of the therapeutic protein when administered  
 CC intravenously to a human. The long term systemic expression results in a  
 CC measurable level of the therapeutic protein being produced in the blood  
 CC of the human for a period of at least 30 days after the administration of  
 CC the RRV vector preparation. RRV's can be used for in vivo delivery of  
 CC therapeutic protein to treat, e.g. haemophilia A, haemophilia B,  
 CC thrombosis, hypercoagulable disorders, liver diseases such as hepatitis,  
 CC disorders such as thalassemia, phenylketonuria, Lesch-Nyhan syndrome,  
 CC severe combined immunodeficiency (SCID), cystic fibrosis, Duchenne's  
 CC Muscular Dystrophy, inherited emphysema, familial hypercholesterolemia,  
 CC diabetes, hypopituitarism, adenine deaminase deficiency, alphas-  
 CC antitrypsin deficiency, Guacher's syndrome, anaemia, infections such as  
 CC HIV infection, high blood pressure, Alzheimer's disease, autoimmune or  
 CC inflammatory disease or graft versus host disease. RRV's are capable of  
 CC surviving inactivation in human serum thereby allowing efficient gene  
 CC transfer over prolonged periods of time.  
 CC (Updated on 25-MAR-2003 to correct PI field.)

XX SQ Sequence 1457 AA;

Query Match 99.9%; Score 7684; DB 19; Length 1457;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 1437; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATRRYLGAVELSWDYMQSDLGELPVDARPPRPVKSPFNTSVVYKKTFLVFETVHLFN 60  
 DB 20 ATRRYLGAVELSWDYMQSDLGELPVDARPPRPVKSPFNTSVVYKKTFLVFETVHLFN 79  
 QY 61 IAKPRPMMGLLGTPTQAEVYDVTVTILKNMASHPVSLHVGVSYWKASGAEYDDQTSQ 120

Db 80 IAKPRPMMGLLGTPTQAEVYDVTVTILKNMASHPVSLHVGVSYWKASGAEYDDQTSQ 139  
 QY 121 REKEDDKVFGGSHHTVQVVLKENGPMASDPLCLTYSYLSHVLDVLDNSGLIGALLVCR 180  
 Db 140 REKEDDKVFGGSHHTVQVVLKENGPMASDPLCLTYSYLSHVLDVLDNSGLIGALLVCR 199  
 QY 181 EGSIAKEKTQTLHKFILLFAVFDGKSWHSETKNSLMQDRDAASARAWPQMTVGVYNR 240  
 Db 200 EGSIAKEKTQTLHKFILLFAVFDGKSWHSETKNSLMQDRDAASARAWPQMTVGVYNR 259  
 QY 241 SLPGLIGCHRSVYVHVGIMGTTPVHSHFLEGHTEFLVRNHRQASLEISITFLTAQTLL 300  
 Db 260 SLPGLIGCHRSVYVHVGIMGTTPVHSHFLEGHTEFLVRNHRQASLEISITFLTAQTLL 319  
 QY 301 MDLGQFLFCHISSHQHDGMEAYVKVDSCEPEEPQLRMKNNEEAEDVDDDLTDSMDVVR 360  
 Db 320 MDLGQFLFCHISSHQHDGMEAYVKVDSCEPEEPQLRMKNNEEAEDVDDDLTDSMDVVR 379  
 QY 361 DDNSPSFIQIRSVAKKHPTWVHYIAAEEEDWDYAPLVAPDDRYSKSYLNNGPORIG 420  
 Db 380 DDNSPSFIQIRSVAKKHPTWVHYIAAEEEDWDYAPLVAPDDRYSKSYLNNGPORIG 439  
 QY 421 RYKVKRWAYTDETFKTRAIQHESGILGPLLYGEVGDTLIIFKNOASRPYNIYPHGI 480  
 Db 440 RYKVKRWAYTDETFKTRAIQHESGILGPLLYGEVGDTLIIFKNOASRPYNIYPHGI 499  
 QY 481 TDVRPLYSRRLPGVKVHLKDFPILPGEIFKYKWTVTVDGPTKSDPRCLTRYTSSFNME 540  
 Db 500 TDVRPLYSRRLPGVKVHLKDFPILPGEIFKYKWTVTVDGPTKSDPRCLTRYTSSFNME 559  
 QY 541 RDLASGLIGPLLCYKESVDQRGNQMSDKRNVILSFVDENRSWYLTENIQRLPNPAG 600  
 Db 560 RDLASGLIGPLLCYKESVDQRGNQMSDKRNVILSFVDENRSWYLTENIQRLPNPAG 619  
 QY 601 VQLEDPEFOASNIWHSINGVYVDSLOLSVCLHVAWYIILSIGAQTDFLSVFSGTYFKH 660  
 Db 620 VQLEDPEFOASNIWHSINGVYVDSLOLSVCLHVAWYIILSIGAQTDFLSVFSGTYFKH 679  
 QY 661 KMVYEDTLTLFPFSGETVFMSENPGWLILGCHNSDFRNRGNTALLKVSSCDNXTGYYE 720  
 Db 680 KMVYEDTLTLFPFSGETVFMSENPGWLILGCHNSDFRNRGNTALLKVSSCDNXTGYYE 739  
 QY 721 DSVEDISAYLLSKNNAIEPRFSQNPVPLKRGHREITRTTLQSDQBEIDYDDTISVEMKK 780  
 Db 740 DSVEDISAYLLSKNNAIEPRFSQNPVPLKRGHREITRTTLQSDQBEIDYDDTISVEMKK 799  
 QY 781 EDFDIYDEBENOSPRSQKTRHYFTAAVERLDYGMSSSPHYLRNRAQSGSYPOPKKV 840  
 Db 800 EDFDIYDEBENOSPRSQKTRHYFTAAVERLDYGMSSSPHYLRNRAQSGSYPOPKKV 859  
 QY 841 FQEFDTGSGFTQPLRYGELNEHLGLLPYIRAEVEDNIMVTFRNOASRPYSFYSLSLYEE 900  
 Db 860 FQEFDTGSGFTQPLRYGELNEHLGLLPYIRAEVEDNIMVTFRNOASRPYSFYSLSLYEE 919  
 QY 901 DQRCQAEPRKNFKVNETKTYFMVQHHMPTKDEFCCKAWAYFSDVLEKOVHSGLIGP 960  
 Db 920 DQRCQAEPRKNFKVNETKTYFMVQHHMPTKDEFCCKAWAYFSDVLEKOVHSGLIGP 979  
 QY 961 LLVCHNTNLNPAHGRQVTVQEFALPFTIFDETKSWYTFENMERNCRAPCNIQMEDTFKE 1020  
 Db 980 LLVCHNTNLNPAHGRQVTVQEFALPFTIFDETKSWYTFENMERNCRAPCNIQMEDTFKE 1039  
 QY 1021 NYRFAHNGYIMDTLPLGVMAQDQRIRWYLLSGNSNENIHSIFSGHVFVTRKKEBKMA 1080  
 Db 1040 NYRFAHNGYIMDTLPLGVMAQDQRIRWYLLSGNSNENIHSIFSGHVFVTRKKEBKMA 1099  
 QY 1081 LYNLYPGVFETVEMLPFSKAGINRVECLIGEHLAGHASTLFLVYSNKCQPLGNASGHIRD 1140  
 Db 1100 LYNLYPGVFETVEMLPFSKAGINRVECLIGEHLAGHASTLFLVYSNKCQPLGNASGHIRD 1159  
 QY 1141 FQITAGSQVQWAPKPLARLHYSGSINAWSTKPEPFSWKVDLLAPMIHGIKTQCARQFS 1200



Db 1160 FQITASQYQWAPKLARLHYSGINAWSTKEPFWIKVDLLAPMIHGIKTOGAROKFS 1219  
 Qy 1201 SLYISQFIIMYSLDGKKQTYRGNSTGTLMVFFGNVDSSGKIKHNFNPPIIARIHLPT 1260  
 Db 1220 SLYISQFIIMYSLDGKKQTYRGNSTGTLMVFFGNVDSSGKIKHNFNPPIIARIHLPT 1279  
 Qy 1261 HYSIRSTRILRMELMCGDLNCSMPLGMSKAISSDAQITASSYFTNMFATWSPSKARLHLQ 1320  
 Db 1280 HYSIRSTRILRMELMCGDLNCSMPLGMSKAISSDAQITASSYFTNMFATWSPSKARLHLQ 1339  
 Qy 1321 RSNARWPOVNNPKWLOVDFOKTMKVGTGVTGQVKSLLTSMYVKEFLISSQDGHQWTLF 1380  
 Db 1340 RSNARWPOVNNPKWLOVDFOKTMKVGTGVTGQVKSLLTSMYVKEFLISSQDGHQWTLF 1399  
 Qy 1381 FONGKVKVQGNQDSFTPVVNSLDPPLTRYLRIRHPSQSWHQIALRMEVLGCEAQDLY 1438  
 Db 1400 FONGKVKVQGNQDSFTPVVNSLDPPLTRYLRIRHPSQSWHQIALRMEVLGCEAQDLY 1457

## RESULT 4

AAW44372  
 ID AAW44372 standard; Protein; 1457 AA.

AC AAW44372;

XX 20-JUL-1998 (first entry)

XX Human Factor VIII SQN deletion mutant.

XX Factor VIII; blood clotting; haemophilia A; gene therapy;

XX retrovirus; vector; human.

XX Homo sapiens.

XX Synthetic.

XX W09800542-A2.

XX 08-JAN-1998.

XX 02-JUL-1997; 97WO-US11785.

XX 04-JUN-1997; 97US-0869309.

XX 03-JUL-1996; 96US-0645601.

XX 13-AUG-1996; 96US-0696381.

XX (CHIR ) CHIRON CORP.

XX Allen JR, Barber JR, Boder M, Chang SM, Chong K;

XX De la Vega D, Depolo NJ, Greengard J, Hsu DC, Ibanez CE;

XX Jolly DJ, Mittelstaedt DM, Prussak CE, Respass JG;

XX WPI; 1998-086967/08.

XX N-PSDB; AAV15338.

XX New replication defective recombinant retroviruses - which express B  
 domain-deleted human factor VIII or human factor IX for the  
 treatment of haemophilia

XX Claim 5; Page 175-180; 236pp; English.

CC This polypeptide comprises the B domain deletion mutant SQN of  
 human Factor VIII. The SQN mutant is created by fusing Ser-743 to  
 Gln-1638 of native Factor VIII (see AAW44373) to form a Ser-Gln-Aan  
 (SQN) link between the A2 and A3 Factor VIII domains. A DNA  
 sequence encoding the SQN deletion mutant is provided in AAV15338.  
 When compared to plasmid-derived Factor VIII, the SQN deletion does  
 not influence the in vivo pharmacokinetics, but the reduced size of  
 the molecule appears to decrease proteolytic degradation. The  
 invention relates to preparations of replication defective  
 recombinant retrovirus (RV) expressing a B domain-deleted human  
 Factor VIII protein, where the recombinant RV is capable of  
 infecting human cells, is resistant to degradation by human  
 complement and is capable of inducing long-term (at least 30 days

CC and up to 6 months or longer post-injection) systemic expression of  
 CC Factor VIII when administered to a haemophilia A patient.

SQ Sequence 1457 AA;

Query Match 99.9%; Score 7684; DB 19; Length 1457;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 1437; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATRRYLGAVELSDVMQSDLGELPVDARPPRPVKSPFNTSVVYKTLFVEFTVHLFN 60

Db 20 ATRRYLGAVELSDVMQSDLGELPVDARPPRPVKSPFNTSVVYKTLFVEFTVHLFN 79

Qy 61 IAKPRPPWMLGLGPTTQAEVYDVITLKNWASHPVSLHAGVSYWKASGAEYDDDTSQ 120

Db 80 IAKPRPPWMLGLGPTTQAEVYDVITLKNWASHPVSLHAGVSYWKASGAEYDDDTSQ 139

Qy 121 REKEDDKVFFGGSHYVQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 180

Db 140 REKEDDKVFFGGSHYVQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 199

Qy 181 EGSIAKEKTQTLHKFILLFAVDEGKSWHSETKNSLMQDRDAASARAWPKMHTVGVYNR 240

Db 200 EGSIAKEKTQTLHKFILLFAVDEGKSWHSETKNSLMQDRDAASARAWPKMHTVGVYNR 259

Qy 241 SLPLGLIGCHRSVYVHWVIGMTTPEVHSIFLEGHTPLVRNHRQASLEISPIITLTAOTLL 300

Db 260 SLPLGLIGCHRSVYVHWVIGMTTPEVHSIFLEGHTPLVRNHRQASLEISPIITLTAOTLL 319

Qy 301 MDLGQFLLCFCHISSHQHDGMEAYVKVDSCEEPQLRMKNNEAEYDDDLTDSMDVVRP 360

Db 320 MDLGQFLLCFCHISSHQHDGMEAYVKVDSCEEPQLRMKNNEAEYDDDLTDSMDVVRP 379

Qy 361 DDNSPSFIQIRSVAKKHPTWVHYTAAEEDWDYAPLVLPDDRYSKSYQLNNGPORIG 420

Db 380 DDNSPSFIQIRSVAKKHPTWVHYTAAEEDWDYAPLVLPDDRYSKSYQLNNGPORIG 439

Qy 421 RYKVKRFMAYTDETFKTREAIQHEGILGLPLYGEGVDTLLIFKNQASRPYNIYPHGI 480

Db 440 RYKVKRFMAYTDETFKTREAIQHEGILGLPLYGEGVDTLLIFKNQASRPYNIYPHGI 499

Qy 481 TDVRPLYSRRLPGVKHLKDFPILPGEIFKYIKWTVTVEDGPTKSDPRCLTRYISSFVWME 540

Db 500 TDVRPLYSRRLPGVKHLKDFPILPGEIFKYIKWTVTVEDGPTKSDPRCLTRYISSFVWME 559

Qy 541 RDLASGLIGPLLI CYKESVDQRGNQIMSDKRNVLFSVFDENRSWYLTENIQRPFPNAG 600

Db 560 RDLASGLIGPLLI CYKESVDQRGNQIMSDKRNVLFSVFDENRSWYLTENIQRPFPNAG 619

Qy 601 VQLEDPEFOASNTMHSINGVYVDSLOLSVCLHEVAYWYILSIGAOTDPLSVFSPGTYFKH 660

Db 620 VQLEDPEFOASNTMHSINGVYVDSLOLSVCLHEVAYWYILSIGAOTDPLSVFSPGTYFKH 679

Qy 661 KMVYEDTLTLFPFSGETVFMSENPGILWILGCHNSDFRNRGMTALLKVSSCDKNTGDIYE 720

Db 680 KMVYEDTLTLFPFSGETVFMSENPGILWILGCHNSDFRNRGMTALLKVSSCDKNTGDIYE 739

Qy 721 DSYEDISAVILLSKNNAIIEPRSPQNPPVLKRQREITRTTLQSDQDEIDYDDTISVEMKK 780

Db 740 DSYEDISAVILLSKNNAIIEPRSPQNPPVLKRQREITRTTLQSDQDEIDYDDTISVEMKK 799

Qy 781 EDPDIDYDEBENSPRSFQKTRHYFTAAVERLWDYCMSSSPHYLRNRAQSGSVPOPKVV 840

Db 800 EDPDIDYDEBENSPRSFQKTRHYFTAAVERLWDYCMSSSPHYLRNRAQSGSVPOPKVV 859

Qy 841 FQEFDTGSGFTQPLYRGELNEHGLGIPYTRAEVEDNMVTFRNOASRPYSFYSSLLSYEE 900

Db 860 FQEFDTGSGFTQPLYRGELNEHGLGIPYTRAEVEDNMVTFRNOASRPYSFYSSLLSYEE 919

Qy 901 DQRQGAEPKRNFKVKNETKTYFMKVOHMAPTKDEPDKAWAYFSDVDLEKDVHSGLIGP 960

Db 920 DQRQGAEPKRNFKVKNETKTYFMKVOHMAPTKDEPDKAWAYFSDVDLEKDVHSGLIGP 979

QY 961 LLVCHTNTLNPAHQGVTVQEPALFTTIDETKSWYFTENMERNCRAPCNQIOWEDPTFK 1020  
Db 980 LLVCHTNTLNPAHQGVTVQEPALFTTIDETKSWYFTENMERNCRAPCNQIOWEDPTFK 1039  
QY 1021 NYRFAHNGIYIMDTLPGLVMAQDQRIWYLLSGNSNENIHSIFSGHVFTVRKCEYKMA 1080  
Db 1040 NYRFAHNGIYIMDTLPGLVMAQDQRIWYLLSGNSNENIHSIFSGHVFTVRKCEYKMA 1099  
QY 1081 LYNLYPGVETVEMLPKSKAGINRVECLIGEHLHAGMSTLFLVYSNKKQTPPLGWSGHIRD 1140  
Db 1100 LYNLYPGVETVEMLPKSKAGINRVECLIGEHLHAGMSTLFLVYSNKKQTPPLGWSGHIRD 1159  
QY 1141 FOITASGOYQWAPKLARLHYSGINAMSTKPFWSIKVDLAPMIHGIKQGAROKFS 1200  
Db 1160 FOITASGOYQWAPKLARLHYSGINAMSTKPFWSIKVDLAPMIHGIKQGAROKFS 1219  
QY 1201 SLVYSOFLIMYSLDGKKWQTYRGNSTGTLWVFGNVDSGIXHNFNPPIIARYIRLHPT 1260  
Db 1220 SLVYSOFLIMYSLDGKKWQTYRGNSTGTLWVFGNVDSGIXHNFNPPIIARYIRLHPT 1279  
QY 1261 HYSIRSTLRMELMGCDLNSCMPLGWSKAISSAQITASSYFTNMPATWSPSKARLHLQ 1320  
Db 1280 HYSIRSTLRMELMGCDLNSCMPLGWSKAISSAQITASSYFTNMPATWSPSKARLHLQ 1339  
QY 1321 RSNARPOVNNKEMLOVDFOKTMKVTGTTQGVKSLLTSMYKVEFLISSQDGHQWTLF 1380  
Db 1340 RSNARPOVNNKEMLOVDFOKTMKVTGTTQGVKSLLTSMYKVEFLISSQDGHQWTLF 1399  
QY 1381 FQNGKVKVQGNQDSFTPVVNSLDPLLRILRIHQSWVHOIALRMEVLGCEAODLY 1438  
Db 1400 FQNGKVKVQGNQDSFTPVVNSLDPLLRILRIHQSWVHOIALRMEVLGCEAODLY 1457

RESULT 5  
AAW23414  
ID AAW23414 standard; Protein; 1471 AA.  
AC AAW23414;  
XX

XX 08-APR-1998 (first entry)  
XX Human B-domain deleted factor VIII protein.

XX Post-translational regulatory element; PRE; enhancer II; intronless gene;  
XX surface antigen gene; cytoplasmic accumulation; targeted delivery;  
XX near consensus splice sequence; blood coagulation factor; factor VIII;  
XX factor IX.

XX Homo sapiens.  
XX  
XX WO9733994-A1.

XX 18-SEP-1997.  
XX 10-MAR-1997; 97WO-US03561.

XX 11-MAR-1996; 96US-0683839.  
XX (IMMU-) IMMUNE RESPONSE CORP.

XX Bidlingmaier S, Ill CR;  
XX WPI; 1997-470874/43.  
XX N-PSDB; AAT73164.

XX Vector for increased expression of intronless genes - comprises  
XX intronless gene with at least one near consensus splice sequence, a  
XX promoter and at least one viral cis-acting post-transcriptional  
XX regulatory element

XX Example 1; Pages 31-36; 59pp; English.

XX The present sequence represents human B-domain deleted factor VIII

CC The cDNA encoding this protein also contains, 3' of the coding region, a  
CC post-translational regulatory element (PRE) of the Hepatitis B virus.  
CC PRE sequences have been shown to function in cis to increase the  
CC steady-state levels of surface gene transcripts by facilitating  
CC cytoplasmic accumulation of these transcripts. The above nucleic acid  
CC sequence is part of a novel vector, comprising an intronless gene  
CC containing 1 or more near consensus splice sequences operably linked to  
CC a promoter sequence so that the gene is transcribed in a cell.  
CC Intronless gene transcripts which contain near consensus splice site  
CC sequences are believed to get tied up in the nucleus of the cell where  
CC splicing occurs, rather than being transported to the cytoplasm where  
CC they can be translated into proteins. The PRE sequences are transcribed  
CC along with the gene, causing export of the gene transcript from the  
CC nucleus into the cytoplasm of the cell. The vector can be used  
CC to increase the expression of an intronless gene containing at least one  
CC near consensus splice sites, preferably cDNA encoding a blood coagulation  
CC factor, particularly Factor VIII or IX. The complex allows the targeted  
CC delivery of the vector to a specific cell, e.g. hepatocytes when the  
CC ligand is an asialoglycoprotein which binds the asialoglycoprotein  
CC receptor present on their surface.

XX Sequence 1471 AA;

Query Match 99.8%; Score 7674; DB 18; Length 1471;  
Besc Local Similarity 99.0%; Pred. No. 0;  
Matches 1438; Conservative 0; Mismatches 0; Indels 14; Gaps 1;

QY 1 ATRRYLGAVELSDWDMQSDIGELPVDARFPFRVPKSPFFNTSVVYKKTFLVEFTVHLFN 60  
Db 20 ATRRYLGAVELSDWDMQSDIGELPVDARFPFRVPKSPFFNTSVVYKKTFLVEFTVHLFN 79  
QY 61 IAKPRPPMGLLGPTIQAEVYDVTVTILKNMASHPSVLHVGVSYSWKASEGAEYDDQTSQ 120  
Db 80 IAKPRPPMGLLGPTIQAEVYDVTVTILKNMASHPSVLHVGVSYSWKASEGAEYDDQTSQ 139  
QY 121 REKEDDKVPGGSHYVQVLKENGPMASDPLCLTYSYLSHVDLVKDLSGLIGALLVCR 180  
Db 140 REKEDDKVPGGSHYVQVLKENGPMASDPLCLTYSYLSHVDLVKDLSGLIGALLVCR 199  
QY 181 EGS�AKEKTQTLHKFILLFAVDEGKSWHSETKNSLMQDRDAASARAPKQHTVNGVYNR 240  
Db 200 EGS�AKEKTQTLHKFILLFAVDEGKSWHSETKNSLMQDRDAASARAPKQHTVNGVYNR 259  
QY 241 SLPLGIGCHRSVYVHWIGMTTPEVHSIFLEGTFLVRNHRQASLEISPIITFLTAOTLL 300  
Db 260 SLPLGIGCHRSVYVHWIGMTTPEVHSIFLEGTFLVRNHRQASLEISPIITFLTAOTLL 319  
QY 301 MDLQGFLLFCHISSHQHDGMEAYVKVDSCPBPQLRMKNEEAEDYDDDLTSEMDVVRP 360  
Db 320 MDLQGFLLFCHISSHQHDGMEAYVKVDSCPBPQLRMKNEEAEDYDDDLTSEMDVVRP 379  
QY 361 DDNSPSFIQIRSVAKKPKTWVHYIAAEEDWDYAPLVLAPODRSYKSOYLNNGPORIG 420  
Db 380 DDNSPSFIQIRSVAKKPKTWVHYIAAEEDWDYAPLVLAPODRSYKSOYLNNGPORIG 439  
QY 421 RYKVKRPMAYTDEFTKTREAIQHSIGILGPLLYGEGVDGTLTLLIPKQASRPYNIYPHGI 480  
Db 440 RYKVKRPMAYTDEFTKTREAIQHSIGILGPLLYGEGVDGTLTLLIPKQASRPYNIYPHGI 499  
QY 481 TDVRPLYSRRLPKGVKHLKDPFLLPGEIFKYKWTVTVEDGDKDPCLRTIYYSFVWME 540  
Db 500 TDVRPLYSRRLPKGVKHLKDPFLLPGEIFKYKWTVTVEDGDKDPCLRTIYYSFVWME 559  
QY 541 RDLASGLIGPLLI CYKESVDQGNQIMSDKRNVLFSVFDNRNRSWYLTENIQRLPNPAG 600  
Db 560 RDLASGLIGPLLI CYKESVDQGNQIMSDKRNVLFSVFDNRNRSWYLTENIQRLPNPAG 619  
QY 601 VQLEDPEFQASNMHSINGYVFDLSQLSVCLHEVAYWYILSIGAQDTFLSVFFSGYTFKH 660  
Db 620 VQLEDPEFQASNMHSINGYVFDLSQLSVCLHEVAYWYILSIGAQDTFLSVFFSGYTFKH 679  
QY 661 KMWVEDTLTLPFPGSETVFMENPGLWILGCHNSDFNRGWTALLKVSSCDKNTGYYE 720

Db 680 KMVEDTLTLFPFSGTVMFMSNPGMLWILGCHNSDFRNGMTALLKVSDDCKNKTGDYYE 739  
QY 721 DSYEDISAYLLSKNNAIEPRFSQN-----PPVLKRHQREITRTTLLQSDQE 766  
Db 740 DSYEDISAYLLSKNNAIEPRFSQNSRHPSTQKQFNATPPVLKRHQREITRTTLLQSDQE 799  
QY 767 EIDYDDTISVEMKEDPDIIDENQSPRQFQKTRHYFIAAVERLWDYQMSSSHVPLVRN 826  
Db 800 EIDYDDTISVEMKEDPDIIDENQSPRQFQKTRHYFIAAVERLWDYQMSSSHVPLVRN 859  
QY 827 RAQSGSVPOPKVVFORFTDGSFTQPLRYRGELNEHLGLGPIYRAEVEDNIMWTFRNOAS 886  
Db 860 RAQSGSVPOPKVVFORFTDGSFTQPLRYRGELNEHLGLGPIYRAEVEDNIMWTFRNOAS 919  
QY 887 RPYSFYSSLSIYBEDORQGAEPKRNFKVKNETKTYFWKHMAPTKDEFDCKAWAYFSD 946  
Db 920 RPYSFYSSLSIYBEDORQGAEPKRNFKVKNETKTYFWKHMAPTKDEFDCKAWAYFSD 979  
QY 947 VDLEKDVHSLIGLPLLVCHTNTLNPAGROVTVQEFALFTTIDETKSWTYFTENMERNCR 1006  
Db 980 VDLEKDVHSLIGLPLLVCHTNTLNPAGROVTVQEFALFTTIDETKSWTYFTENMERNCR 1039  
QY 1007 APCNIQMEDPTFKENRFHAINGYIMDTLPLGVMAQDORIRWYLLSMGNSNENIHSIFSG 1066  
Db 1040 APCNIQMEDPTFKENRFHAINGYIMDTLPLGVMAQDORIRWYLLSMGNSNENIHSIFSG 1099  
QY 1067 HVFTVRKKEBYKMAVNLNYPGVFETVEMLPSPKAGINRVECLIGEHLHAGNSTLFLVYSNK 1126  
Db 1100 HVFTVRKKEBYKMAVNLNYPGVFETVEMLPSPKAGINRVECLIGEHLHAGNSTLFLVYSNK 1159  
QY 1127 CQPLGMASGHIRDFOITASQYQGWAPKAPLARLHYSGINAWSTKEPFSWIKVDLLAPMI 1186  
Db 1160 CQPLGMASGHIRDFOITASQYQGWAPKAPLARLHYSGINAWSTKEPFSWIKVDLLAPMI 1219  
QY 1187 IHGKTKGAKQKSSLYISQFIIMYSLDGKKWQTYRGNSTGLMVRFGNVDSSGIKHNI 1246  
Db 1220 IHGKTKGAKQKSSLYISQFIIMYSLDGKKWQTYRGNSTGLMVRFGNVDSSGIKHNI 1279  
QY 1247 NPPIIARYIRLHPTHYSIRTLRMELMGCDLNSCMPLGMESKAISDAQITASSYFTNMF 1306  
Db 1280 NPPIIARYIRLHPTHYSIRTLRMELMGCDLNSCMPLGMESKAISDAQITASSYFTNMF 1339  
QY 1307 ATWSPSKARLHLQGRSNAMPQVNNPKEWLQVDFQKTMKVTVGTQGVKSLLTSMYKKEF 1366  
Db 1340 ATWSPSKARLHLQGRSNAMPQVNNPKEWLQVDFQKTMKVTVGTQGVKSLLTSMYKKEF 1399  
QY 1367 LISSQDGHQWTLFFQNGKVKVFGNQDSTFPVNSLDPLLTLYLRIHQPQSWHQTALR 1426  
Db 1400 LISSQDGHQWTLFFQNGKVKVFGNQDSTFPVNSLDPLLTLYLRIHQPQSWHQTALR 1459  
QY 1427 MEVLGCEAODLY 1438  
Db 1460 MEVLGCEAODLY 1471

## RESULT 6

AAB67959  
ID AAB67959 standard; Protein; 1471 AA.

XX XX

AC AAB67959;

XX XX

DT 29-JUN-2001 (first entry)

XX Amino acid sequence of human B-domain deleted factor VIII.

XX Adeno-associated virus vector; B-domain; factor VIII; haemophilia A;  
coagulation disorder.

XX Homo sapiens.

XX W0200127303-A1.

XX 19-APR-2001.

PD

XX 12-OCT-2000; 2000WO-US28221.  
XX 12-OCT-1999; 99US-0158780.  
XX (UYNC-) UNIV NORTH CAROLINA.  
XX Walsh CE, Chao H, Burstein H, Lynch CM, Stepan AM, Munson K;  
PI WPI; 2001-273781/28.  
XX N-PSDB; AAR64647.  
DR  
XX New recombinant adeno-associated virus vector, useful for treating  
PT haemophilia A, comprises heterologous nucleotide sequence encoding  
PT B-domain deleted human factor VIII operably linked with liver-preferred  
PT expression control element -  
XX  
XX Example 1; Fig 1; 87pp; English.  
XX The specification describes a recombinant adeno-associated virus (rAAV)  
CC vector. The vector comprises a heterologous nucleotide sequence  
CC encoding B-domain deleted factor VIII operably linked with at least one  
CC enhancer and at least one promoter. The method results in the production  
CC of high titer rAAV vector stocks carrying the B-domain deleted factor  
CC VIII transgenes and expression cassettes, which generate adequate titers  
CC of virus for in vivo administration. The recombinant vectors are useful  
CC for treating haemophilia A, where the liver expresses the encoded  
CC B-domain deleted factor VIII, which is secreted into the blood. They are  
CC also useful for the treatment of other coagulation disorders. The  
CC present sequence represents a B-domain deleted factor VIII.  
XX  
XX Sequence 1471 AA;  
SQ  
Query Match 99.8%; Score 7674; DB 22; Length 1471;  
Best Local Similarity 99.0%; Pred. No. 0;  
Matches 1438; Conservative 0; Mismatches 0; Indels 14; Gaps 1;  
QY 1 ATRRYLGAVELSWDMQSDLGELPVDARPPRPVKSPFPNTSVYKTLFVEFTVHLFN 60  
Db 20 ATRRYLGAVELSWDMQSDLGELPVDARPPRPVKSPFPNTSVYKTLFVEFTVHLFN 79  
QY 61 IAKPRPFWMLLGPFTTQAEVYDVTWITLKNMASHPVSLHAGVSYNKASGAYDDQTSQ 120  
Db 80 IAKPRPFWMLLGPFTTQAEVYDVTWITLKNMASHPVSLHAGVSYNKASGAYDDQTSQ 139  
QY 121 REKDDKVPFGSGSHYVWVLKENGPMASDPLCLTYSLSHVDLVKDLNSGLLVCVR 180  
Db 140 REKDDKVPFGSGSHYVWVLKENGPMASDPLCLTYSLSHVDLVKDLNSGLLVCVR 199  
QY 181 EGSIAKEKTQTLHKFTILLFAVFDGKSWHSETKNSLMQDRDAASARAWPKMHTVNGVNR 240  
Db 200 EGSIAKEKTQTLHKFTILLFAVFDGKSWHSETKNSLMQDRDAASARAWPKMHTVNGVNR 259  
QY 241 SLPGLIGCHRKSVYMHVIGMTTPEVHSIFLEGHTFLVRNHRQASLEISPTITLTAQTL 300  
Db 260 SLPGLIGCHRKSVYMHVIGMTTPEVHSIFLEGHTFLVRNHRQASLEISPTITLTAQTL 319  
QY 301 MDLGQFLFCHISSHQHGMVAVKVDSCPEEPQLRMKNNEAEYDDDLTDSEMDVVR 360  
Db 320 MDLGQFLFCHISSHQHGMVAVKVDSCPEEPQLRMKNNEAEYDDDLTDSEMDVVR 379  
QY 361 DDNSPSFQIRSVAKKHPTWVHYIAABEEDWDYAPLVLPDDRYSKSOYLANGPQRIG 420  
Db 380 DDNSPSFQIRSVAKKHPTWVHYIAABEEDWDYAPLVLPDDRYSKSOYLANGPQRIG 439  
QY 421 RYKVKRFWAYTDETPKTREAIQHESGILGPLYGVGDTLLIFPKQASRPNIYPHGI 480  
Db 440 RYKVKRFWAYTDETPKTREAIQHESGILGPLYGVGDTLLIFPKQASRPNIYPHGI 499  
QY 481 TDVRPLYSRRLPKGVKHLKDFPILPGEIIFYKWTVTVEGPTKSDPRCLTRYSSFFNME 540  
Db 500 TDVRPLYSRRLPKGVKHLKDFPILPGEIIFYKWTVTVEGPTKSDPRCLTRYSSFFNME 559

QY 541 RDLASGLIGPLLICVKEVDQGNQIMSDKKNVILFSVFDENRSHWLTENTQRELPNDAG 600  
Db 560 RDLASGLIGPLLICVKEVDQGNQIMSDKKNVILFSVFDENRSHWLTENTQRELPNDAG 619  
QY 601 VOLEDPEFQASNIHMSINGYVFDLSQLSVCLHEVAYWILSGTDFLSVFFSGYTFKH 660  
Db 620 VOLEDPEFQASNIHMSINGYVFDLSQLSVCLHEVAYWILSGTDFLSVFFSGYTFKH 679  
QY 661 KMVYEDTLTLFPFSGETVFMENPGLMILGCHNSDFRNQMTALLKVSSCDKNTGDIYE 720  
Db 680 KMVYEDTLTLFPFSGETVFMENPGLMILGCHNSDFRNQMTALLKVSSCDKNTGDIYE 739  
QY 721 DSYEDISAYLISKNAIIEPRFSQN-----PPVLKHQREIIRTTLOSQOE 766  
Db 740 DSYEDISAYLISKNAIIEPRFSQNSRHPSTKQFNATPPVLKHQREIIRTTLOSQOE 799  
QY 767 EIDYDDTISVEMKEDFDIYDEENQSPRSFQKTRHYFIAAVERLWDYGMSSPHVLRN 826  
Db 800 EIDYDDTISVEMKEDFDIYDEENQSPRSFQKTRHYFIAAVERLWDYGMSSPHVLRN 859  
QY 827 RAQSGSVFPQKVVQFETDGSFTQPLRYGELNEHLGLGPGYIRAEVEDNIMVTFRNQAS 886  
Db 860 RAQSGSVFPQKVVQFETDGSFTQPLRYGELNEHLGLGPGYIRAEVEDNIMVTFRNQAS 919  
QY 887 RPYSPYSLISVEEDQROGAPRKNFVQNETKTYFWKVQHHMPTKDEPCKAWAYPSD 946  
Db 920 RPYSPYSLISVEEDQROGAPRKNFVQNETKTYFWKVQHHMPTKDEPCKAWAYPSD 979  
QY 947 VDLEKDVHSGLIGPLLVCHTNTLPAHGRQVTVQEPALFFTFIDETKSNVYFENNERCR 1006  
Db 980 VDLEKDVHSGLIGPLLVCHTNTLPAHGRQVTVQEPALFFTFIDETKSNVYFENNERCR 1039  
QY 1007 APCNQIMQEDPFPKYNRFAHNGYIMDTLPGLVMAQDQIRIRWYLLSMGNSNENIHSIFSG 1066  
Db 1040 APCNQIMQEDPFPKYNRFAHNGYIMDTLPGLVMAQDQIRIRWYLLSMGNSNENIHSIFSG 1099  
QY 1067 HVFTVRKKEEYKMALYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTFLVYGNK 1126  
Db 1100 HVFTVRKKEEYKMALYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTFLVYGNK 1159  
QY 1127 COTPLGMSAGHIRDFOITASGOYGQWAPKLAHLHYSGSINASTKEPFSWIKVDLLPMI 1186  
Db 1160 COTPLGMSAGHIRDFOITASGOYGQWAPKLAHLHYSGSINASTKEPFSWIKVDLLPMI 1219  
QY 1187 IHGIKTQARQKXFSLLXISQFIIMYSLDGKKWOTYRGNSTGTLMVFFGNVDSSGKKNIF 1246  
Db 1220 IHGIKTQARQKXFSLLXISQFIIMYSLDGKKWOTYRGNSTGTLMVFFGNVDSSGKKNIF 1279  
QY 1247 NPPIIARIYRLHPTHYSIRSTRMELMGCDLNSCMPLGMSKALSDAQITASSYFTNMF 1306  
Db 1280 NPPIIARIYRLHPTHYSIRSTRMELMGCDLNSCMPLGMSKALSDAQITASSYFTNMF 1339  
QY 1307 ATWSPSKARLHLQGRSNARWPOVNNPKWLQYDFOKTMKVTGVTGQVKSLTSMYVKEF 1366  
Db 1340 ATWSPSKARLHLQGRSNARWPOVNNPKWLQYDFOKTMKVTGVTGQVKSLTSMYVKEF 1399  
QY 1367 LISSSQDGHQWTLFPQNGKVKVFCQGNQDSFTPVVNSLDPPLLTRYLRHPSWVHQIALR 1426  
Db 1400 LISSSQDGHQWTLFPQNGKVKVFCQGNQDSFTPVVNSLDPPLLTRYLRHPSWVHQIALR 1459  
QY 1427 MEVLGCEAQDLY 1438  
Db 1460 MEVLGCEAQDLY 1471

## RESULT 7

AAR12971  
ID AAR12971 standard; protein; 1440 AA.  
XX AC  
AC AAR12971;  
XX 25-MAR-2003 (updated)  
DT 09-JAN-2003 (updated)

Query Match 99.6%; Score 7659.5; DB 12; Length 1440;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 1434; Conservative 1; Mismatches 3; Indels 1; Gaps 1;  
QY 1 ATRRYYLGAVELSWDMQSDLGELPVDARPPRPVKSPFPNTSVVYKKTLEFVETVHLFN 60  
Db 2 ATRRYYLGAVELSWDMQSDLGELPVDARPPRPVKSPFPNTSVVYKKTLEFVETVHLFN 61  
QY 61 IAKPRPPMMGLGPTIQAEVYDVTVTILKNNAHPVSLHAYGVSYWKASGAEYDDQTSQ 120  
Db 62 IAKPRPPMMGLGPTIQAEVYDVTVTILKNNAHPVSLHAYGVSYWKASGAEYDDQTSQ 121  
QY 121 REKEDDKVFPQSGSHYVQVLKENGPMASDPLCTYTSYLSHVDLVKDLNGLIGALLVCR 180

DT 02-OCT-1991 (first entry)  
XX Factor VIII:SQ.  
XX Factor VIII; B domain; haemophilia.  
XX Unidentified.  
XX WO9109122-A.  
XX 27-JUN-1991.  
XX 06-DEC-1990; 90WO-SE00809.  
XX 15-DEC-1989; 89SE-0004239.  
XX (KABI ) KABIVITRUM AB.  
XX Almstedt AB, Hellstrom EM, Larsson K, Lind P, Sandberg HI;  
XX Spira J, Sydowback MM;  
XX WPI; 1991-208148/28.  
XX Recombinant human factor VIII deriv. deoxyribonucleic acid -  
XX encoding protein comprising two chains linked by segment of B domain.  
XX Disclosure; Fig 1; 35pp; English.

The protein is a fusion between Phe 742 and Ser 1637 of the factor VIII protein (factor VIII:SQ). In order to produce a factor VIII deletion derivative that can be produced in vivo and/or in vitro, to a two chain protein consisting of polypeptide chains of 90 kD and 80 kD, the amino acid sequences surrounding Arg 740 and Arg 1648 have to be conserved in order to preserve the structural requirements for correct cleavage. In this example, amino acids 743 to 1636 of the full-length factor VIII polypeptide are deleted. A new polypeptide chain is obtained where there are 14 amino acids linking Arg 740 and Arg 1648. Of these 14 amino acids, the sequence of the five N-terminal ones directly corresponds to the five amino acids following Arg 740 in full-length factor VIII. Also, the sequence of the 12 C-terminal amino acids of the above 14 amino acids fragment directly corresponds to the 12 amino acids preceding Glu 1689 in full-length factor VIII, thus creating a 3 amino acid overlap between the N- and C-terminal regions of the B-domain.

The factor VIII deriv. is useful for treating haemophilia or haemophilia A. It has the biological characteristics of plasma derived factor VIII. In order to index this example, the factor VIII:QD amino acid sequence was retrieved from WO8800831 (AAP80265). The amino acid numbering in the above comments is reproduced from the fig. description in the specification. Note that Arg 740 is Arg 742 in the indexed sequence, etc. Also, Asn 745 (N-terminal link overlap) is Asp 747 in AAP80265, but indexed as Asn to reproduce the fusion fragment as shown in fig.1, and Asn 745 (C-terminal link overlap). (Updated on 09-JAN-2003 to add missing OS field.) (Updated on 25-MAR-2003 to correct PI field.)

Sequence 1440 AA;

Db 122 REKEDKVPFGSGSHYVQVWKENGPMASDPLCLTYSYLSHVLDVVDLNSGLIGALLVCR 181  
QY 181 EGS�AKEKTQTLHKFILLPAVFPDEGKSWHSETKNSLMQDRDAASARAWPMQHTVNGVNR 240  
Db 182 EGS�AKEKTHTLHKFILLPAVFPDEGKSWHSETKNSLMQDRDAASARAWPMQHTVNGVNR 241  
QY 241 S-LPGLIGCHRKSVYWHVIGMGTTPPEVHSIFLGHFTFLVNRHQASLEISPIIFLTAQTL 299  
Db 242 SLPLGLIGCHRKSVYWHVIGMGTTPPEVHSIFLGHFTFLVNRHQASLEISPIIFLTAQTL 301  
QY 300 LMDLQGLFLFCHLSSHQHDGMEAYVKVDSCEPSPQLRMKNNEAEYDDDLTDTSEMDVVR 359  
Db 302 LMDLQGLFLFCHLSSHQHDGMEAYVKVDSCEPSPQLRMKNNEAEYDDDLTDTSEMDLVR 361  
QY 360 FDDNDSPSFIQIRSAVAKHKPTWHYIAAEEEDWDYAPLVLAPEDDRSYKSQYLNNGPORI 419  
Db 362 FDDNDSPSFIQIRSAVAKHKPTWHYIAAEEEDWDYAPLVLAPEDDRSYKSQYLNNGPORI 421  
QY 420 GRKYKVRPMAYTDETFKTREATOHESGILGPLLYGEVGDITLLIIFKNQASRPYNIYPHG 479  
Db 422 GRKYKVRPMAYTDETFKTREATOHESGILGPLLYGEVGDITLLIIFKNQASRPYNIYPHG 481  
QY 480 ITDVRPLYSRRLPGVKHLDKDFILPGEIPKYKWTVTVEDGPTKSDPRCLTRYVSSFVNM 539  
Db 482 ITDVRPLYSRRLPGVKHLDKDFILPGEIPKYKWTVTVEDGPTKSDPRCLTRYVSSFVNM 541  
QY 540 ERDLASGLIGPLLI CYKESVDQRGNQIMSDKRNVLFSVFDENRSWYLTENIQRFLEPNA 599  
Db 542 ERDLASGLIGPLLI CYKESVDQRGNQIMSDKRNVLFSVFDENRSWYLTENIQRFLEPNA 601  
QY 600 GVQLEDEFOASNIMHSINGVYFDSLSQLSVCLHEVAYWYILSTGAQTDLFVSFFSGYTFK 659  
Db 602 GVQLEDEFOASNIMHSINGVYFDSLSQLSVCLHEVAYWYILSTGAQTDLFVSFFSGYTFK 661  
QY 660 HKMAYEDTLTLFPSETVPMSENGLWILGCHNSDFRNGMTALLKYSSCDKNTGDYY 719  
Db 662 HKMAYEDTLTLFPSETVPMSENGLWILGCHNSDFRNGMTALLKYSSCDKNTGDYY 721  
QY 720 EDSYEDIAYLLSKNNAIEPRSPQNPVLKRQREITRTLQSDQEEIDYDITISVEMK 779  
Db 722 EDSYEDIAYLLSKNNAIEPRSPQNPVLKRQREITRTLQSDQEEIDYDITISVEMK 781  
QY 780 KEDFDIYDEBENQSPRSFQKTRHYFTIAAVERLWDYGMSSSPHVLNRQAQSGVPOFKKY 839  
Db 782 KEDFDIYDEBENQSPRSFQKTRHYFTIAAVERLWDYGMSSSPHVLNRQAQSGVPOFKKY 841  
QY 840 VFQEFDTGSGFTQPLYRGEINEHLGILGPYIRAEVEDNIMVTFRNQASRPYSFYSSLSIYE 899  
Db 842 VFQEFDTGSGFTQPLYRGEINEHLGILGPYIRAEVEDNIMVTFRNQASRPYSFYSSLSIYE 901  
QY 900 EDORQAEPRKNFVKNETKYFWKQHHMPTKDEDFCKAWAYFSDVDLEKDVHSGLIG 959  
Db 902 EDORQAEPRKNFVKNETKYFWKQHHMPTKDEDFCKAWAYFSDVDLEKDVHSGLIG 961  
QY 960 PLLVCHTNTLPAHGQVTVQEFALPFTIPDETQSWYFTENMERNCRAPCNIQMEDPTFK 1019  
Db 962 PLLVCHTNTLPAHGQVTVQEFALPFTIPDETQSWYFTENMERNCRAPCNIQMEDPTFK 1021  
QY 1020 ENYRFAINGYIMDTPLGLVMAQDQIRWYLLSMGNSNENIHSIFPSGHVFTVRKKEEYKM 1079  
Db 1022 ENYRFAINGYIMDTPLGLVMAQDQIRWYLLSMGNSNENIHSIFPSGHVFTVRKKEEYKM 1081  
QY 1080 ALYNLPVGVETVEMLPSPKAGIWRVECLIGEHLHAGMSTLFLVYNNKQCPPLGMAQSHIR 1139  
Db 1082 ALYNLPVGVETVEMLPSPKAGIWRVECLIGEHLHAGMSTLFLVYNNKQCPPLGMAQSHIR 1141  
QY 1140 DFOITASQYGGQWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMIIGHIKTQGARQKF 1199  
Db 1142 DFOITASQYGGQWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMIIGHIKTQGARQKF 1201  
QY 1200 SSLYISQFIIMYSLDGKQWQTYRGNSTGTLMVFFGNVDSSGIKNIFNPPIIARYIRLHP 1259  
Db 1202 SSLYISQFIIMYSLDGKQWQTYRGNSTGTLMVFFGNVDSSGIKNIFNPPIIARYIRLHP 1261

QY 1260 THYSIRSTLRMELMGCDLNSCSMPLGMSKASDAQITASSYFTNNMFATWSPSKARLHLQ 1319  
Db 1262 THYSIRSTLRMELMGCDLNSCSMPLGMSKASDAQITASSYFTNNMFATWSPSKARLHLQ 1321  
QY 1320 GRSNARVPQNNPKWELQVDFOKTKWKTGVTGTTQGVKSLLTSMYVKEFLISSODGHQWTL 1379  
Db 1322 GRSNARVPQNNPKWELQVDFOKTKWKTGVTGTTQGVKSLLTSMYVKEFLISSODGHQWTL 1381  
QY 1380 FFONGKVKVQGNQDSFTPVVNSLDPPLLTRYLRIHPQSWHQIALRMEVLGCEAODLY 1438  
Db 1382 FFONGKVKVQGNQDSFTPVVNSLDPPLLTRYLRIHPQSWHQIALRMEVLGCEAODLY 1440

## RESULT 8

AAE10832  
ID AAE10832 standard; Protein; 1459 AA.  
XX  
AC AAE10832;  
XX  
DT 18-DEC-2001 (first entry)  
XX Human factor VIII mutein encoded by vector pTGF8-2hyg-s.  
DE Human; haemostatic; coagulant; blood clotting factor; factor VIII;  
XX factor IX; therapy; haemophilia A; mutant; mutein.  
XX Homo sapiens.  
OS Synthetic.  
XX  
FH Key  
FT Peptide 1..19  
FT Protein /label= Signal\_peptide  
FT Protein 20..1459 /product= "Human mature factor VIII mutein"  
FT Region 780..775 /label= Linker-peptide  
FT  
XX W0200170968-A2.  
XX  
XX  
PD 27-SEP-2001.  
XX  
PF 21-MAR-2001; 2001WO-EP03220.  
XX  
XX 22-MAR-2000; 2000EP-0106225.  
PR 08-MAY-2000; 2000US-203249P.  
XX  
XX (OCTA-) OCTAGENE GMBH.  
XX  
PI Hauser C, Hoerster A, Schroeder C, Lehnerer M;  
XX  
DR WPI; 2001-590175/66.  
DR N-PSDB; AAD18176.  
XX  
PT Recombinantly producing human blood coagulation factors VIII and IX for  
XX use in treating hemophilia -  
PS Claim 17; Page 86-90; 104pp; English.  
XX

CC The present invention relates to an improved method for the production  
CC of recombinant human blood clotting factors, especially factors VIII and  
CC IX, using an immortalised cell line stably expressing viral transcription  
CC activators and carrying a vector comprising a promoter and a sequence  
CC encoding the blood coagulation factor. The factor VIII mutein or a gene  
CC transfer vector is used in the preparation of agents for treating  
CC haemophilia, especially haemophilia A. The present sequence is  
CC human factor VIII mutein encoded by vector pTGF8-2hyg-s. The vector  
CC contains a silent mutation, resulting in a factor VIII mutein having  
CC the substitution of the B-domain of wild-type factor VIII by the  
CC linker peptide.

Sequence 1459 AA;

Query Match				99.2%;	Score 7626;	DB 22;	Length 1459;
Best Local Similarity				99.1%;	Pred. No. 0;		
Matches 1431;				Conservative 0;	Mismatches 3;	Indels 10;	Gaps 2;
QY	1	ATRRYYLGAVELSWDMQSDLGELPVDARPPRPVKSPFPNTSVVYKKTLPVFETVHLFN	60				
Db	20	ATRRYYLGAVELSWDMQSDLGELPVDARPPRPVKSPFPNTSVVYKKTLPVFETVHLFN	79				
QY	61	IAKPPPPWMLGPTTQAEVYDVTVITLKNMASHPVSLHAGVSYMKASGEAYDDQTSQ	120				
Db	80	IAKPPPPWMLGPTTQAEVYDVTVITLKNMASHPVSLHAGVSYMKASGEAYDDQTSQ	139				
QY	121	REKEDDKVPFGGSHYVQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR	180				
Db	140	REKEDDKVPFGGSHYVQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR	199				
QY	181	EGSLAKEKTQTLHKETILLFAVDEGKSWHSETKNSLMQDRDAASARAPKMHVTNGVYNR	240				
Db	200	EGSLAKEKTQTLHKETILLFAVDEGKSWHSETKNSLMQDRDAASARAPKMHVTNGVYNR	259				
QY	241	SLPGLIGCHRSVYVHVIQMGTTPEVHSIFLEGHTFLVRNHRQASLEISPTIFLTAQTLL	300				
Db	260	SLPGLIGCHRSVYVHVIQMGTTPEVHSIFLEGHTFLVRNHRQASLEISPTIFLTAQTLL	319				
QY	301	MDLGQFLFCHLSSHQDGMAYKYVDSCEEPQOLRMKNNEBEADYDDDLTDSEMDVYRF	360				
Db	320	MDLGQFLFCHLSSHQDGMAYKYVDSCEEPQOLRMKNNEBEADYDDDLTDSEMDVYRF	379				
QY	361	DDNSPSFTQIRSVAKKHPTWVHYIAAEEEDWDYAPLVAPDDRSYKSYQLNNGPQRIQ	420				
Db	380	DDNSPSFTQIRSVAKKHPTWVHYIAAEEEDWDYAPLVAPDDRSYKSYQLNNGPQRIQ	439				
QY	421	RKYKVRFMAYTDETFKTREAIQHESGILGPLLYGEVGTDLIIIFKNOASRPNTIYPHGI	480				
Db	440	RKYKVRFMAYTDETFKTREAIQHESGILGPLLYGEVGTDLIIIFKNOASRPNTIYPHGI	499				
QY	481	TDVRLPYRRLPKGVKHLKDFILPGEIPKYKWTVTVEDGPTKSDPRCLTRYYSFVNM	540				
Db	500	TDVRLPYRRLPKGVKHLKDFILPGEIPKYKWTVTVEDGPTKSDPRCLTRYYSFVNM	559				
QY	541	RLASGLIGPLLI CYKESVDQGNQIMSDKRNVLFSVFDENRSMYLTENIQRLPNPAG	600				
Db	560	RLASGLIGPLLI CYKESVDQGNQIMSDKRNVLFSVFDENRSMYLTENIQRLPNPAG	619				
QY	601	VOLEDPFOASNMHSINGVYFDSLQJLVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKH	660				
Db	620	VOLEDPFOASNMHSINGVYFDSLQJLVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKH	679				
QY	661	KMVEEDTLTLPFSGETVFMSENPGILWILGCHNSDFNRGMTALLKVSSCDKNTGDYVE	720				
Db	680	KMVEEDTLTLPFSGETVFMSENPGILWILGCHNSDFNRGMTALLKVSSCDKNTGDYVE	739				
QY	721	DSYEDISAYLLSKNNAIBSPSQNPVPLKRROR-----EITFTTLQSDQEEIDYDDTI	774				
Db	740	DSYEDISAYLLSKNNAIBSPSQN-----SRHQAYRYRGEITRTTLQSDQEEIDYDDTI	795				
QY	775	SVEMKEDFDIYDEDEMQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNRNAQSGVP	834				
Db	796	SVEMKEDFDIYDEDEMQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNRNAQSGVP	855				
QY	835	QFKVVQFBFTGSGTQPLRYGELNEHLGLLGPYIRAEVEDNIMVTFRNOASRPYSFYSS	894				
Db	856	QFKVVQFBFTGSGTQPLRYGELNEHLGLLGPYIRAEVEDNIMVTFRNOASRPYSFYSS	915				
QY	895	LISYEDORQGAEPKRNKFNKPYKVOHMAPTKDEPDCWAYFSVDVLEKDVH	954				
Db	916	LISYEDORQGAEPKRNKFNKPYKVOHMAPTKDEPDCWAYFSVDVLEKDVH	975				
QY	955	SGLIGPLLYCHTNTLMPAHGRQVTVQEFALFTTIFDETSKSWYFTENMERNCRAPCNIQME	1014				
Db	976	SGLIGPLLYCHTNTLMPAHGRQVTVQEFALFTTIFDETSKSWYFTENMERNCRAPCNIQME	1035				
QY	1015	DPTFKENYRFAHNGYIMDTLPGLVMAQDQRIRWYLLSMGSENHSHIFSGHVTVRKK	1074				

Db	1036	DPTFKENYRFAHNGYIMDTLPGLVMAQDQRIRWYLLSMGSENHSHIFSGHVTVRKK	1095
QY	1075	EYKXALYNLYPGVFETVEMLPKSGIWRVECLIGEHLHAGMSTFLVYSNKCOTPLGMA	1134
Db	1096	EYKXALYNLYPGVFETVEMLPKSGIWRVECLIGEHLHAGMSTFLVYSNKCOTPLGMA	1155
QY	1135	SGHIRDFQITASGOYQOWAPKPLARLHYSGSINAWSTKEPFSWIKVDLLAPMI IHGKTQ	1194
Db	1156	SGHIRDFQITASGOYQOWAPKPLARLHYSGSINAWSTKEPFSWIKVDLLAPMI IHGKTQ	1215
QY	1195	ARQFSSLYISQFIIMYSLDGKWKQTYRGNSTGTLMVFFGNVDSSGIGHNIFNPPIIARY	1254
Db	1216	ARQFSSLYISQFIIMYSLDGKWKQTYRGNSTGTLMVFFGNVDSSGIGHNIFNPPIIARY	1275
QY	1255	IRLHPTHYSIRSTRLMELMGCDLNSCSMPLGMSKASDAQITASSYFTNMFATWSPSKA	1314
Db	1276	IRLHPTHYSIRSTRLMELMGCDLNSCSMPLGMSKASDAQITASSYFTNMFATWSPSKA	1335
QY	1315	RLHLQGRSNARPOVNNPKWEHLQVDFOKTMKVTGTTQGVKSLLTSMYVKEFLISSQDG	1374
Db	1336	RLHLQGRSNARPOVNNPKWEHLQVDFOKTMKVTGTTQGVKSLLTSMYVKEFLISSQDG	1395
QY	1375	HOWTLFFQNGKVKVFGNQDSFTPVVNSLDPEPLLTRYLRIHPQSWHQAIALRMEVLGCEA	1434
Db	1396	HOWTLFFQNGKVKVFGNQDSFTPVVNSLDPEPLLTRYLRIHPQSWHQAIALRMEVLGCEA	1455
QY	1435	QDLY 1438	
Db	1456	QDLY 1459	
RESULT 9			
AAE10833			
ID	AAE10833	standard; Protein; 1459 AA.	
XX	AAE10833;		
XX	18-DEC-2001	(first entry)	
XX	Human factor VIII mutein encoded by vector pTGF8-3.		
XX	Human; haemostatic; coagulant; blood clotting factor; factor VIII;		
XX	factor IX; therapy; haemophilia A; mutant; mutein.		
XX	Homo sapiens.		
XX	Synthetic.		
XX	Key	Location/Qualifiers	
XX	Peptide	1..19	
XX	Protein	/label= Signal peptide	
XX		20..1459	
XX		/product= "Human mature factor VIII mutein"	
XX	Misc-difference	181	
XX		/note= "Wild type Val substituted with Ala"	
XX	Region	760..775	
XX		/label= Linker-peptide	
XX	Misc-difference	1350	
XX		/note= "Wild type Val substituted with Glu"	
XX	WO200170968-A2.		
XX	27-SEP-2001.		
XX	21-MAR-2001; 2001WO-EP03220.		
XX	22-MAR-2000; 2000EP-0106225.		
XX	08-MAY-2000; 2000US-203249E.		
XX	(OCTA-) OCTAGENE GMBH.		
XX	Hauser C, Hoerster A, Schroeder C, Lehnerer M;		

DR WPI; 2001-590175/66.  
XX N-PSDB; AAD18176.  
PT Recombinantly producing human blood coagulation factors VIII and IX for  
PT use in treating hemophilia -  
XX Claim 17; Page 99-103; 104pp; English.  
XX  
CC The present invention relates to an improved method for the production  
CC of recombinant human blood clotting factors, especially factors VIII and  
CC IX, using an immortalised cell line stably expressing viral transcription  
CC activators and carrying a vector comprising a promoter and a sequence  
CC encoding the blood coagulation factor. The factor VIII mutin or a gene  
CC transfer vector is used in the preparation of agents for treating  
CC haemophilia, especially haemophilia A. The present sequence is human  
CC factor VIII mutin encoded by vector pTGF8-3. This sequence contains  
CC 2 mutations and the B-domain of wild-type factor VIII is substituted  
CC with a linker peptide.  
XX  
SQ Sequence 1459 AA;

Query Match 99.0%; Score 7616; DB 22; Length 1459;  
Best Local Similarity 99.0%; Pred. No. 0;  
Matches 1429; Conservative 0; Mismatches 5; Indels 10; Gaps 2;

QY 1 ATRRYLGAVELSWDMQSDLGELPVDAREPPRPVPSKSPFNTSVYKTLFVEFTVHLEN 60  
DB 20 ATRRYLGAVELSWDMQSDLGELPVDAREPPRPVPSKSPFNTSVYKTLFVEFTVHLEN 79  
QY 61 IAKPRPPWMLLOPTQAEVYDVTVTTLKMAHPVSLHAGVSYKASGAEYDQTSQ 120  
DB 80 IAKPRPPWMLLOPTQAEVYDVTVTTLKMAHPVSLHAGVSYKASGAEYDQTSQ 139  
QY 121 REKEDKVPFGSGHTYVQVLYKNGPMASDPLCLTYSLSHVDLVKDLNSGLIGALLVCR 180  
DB 140 REKEDKVPFGSGHTYVQVLYKNGPMASDPLCLTYSLSHVDLVKDLNSGLIGALLVCR 199  
QY 181 EGS LAKEKTLHLKFIILFAVFDGKSWHSETKNSLMQDRDAASARAPKMHVTVNGVNR 240  
DB 200 EGS LAKEKTLHLKFIILFAVFDGKSWHSETKNSLMQDRDAASARAPKMHVTVNGVNR 259  
QY 241 SLPLGLGCHRSYVHVHVGMTTPEVHSIFLEGHTFLVRNHRQASLEISPIFLTAQTLL 300  
DB 260 SLPLGLGCHRSYVHVHVGMTTPEVHSIFLEGHTFLVRNHRQASLEISPIFLTAQTLL 319  
QY 301 MDLGOFLFCHISHOHDMGEAYVKVDSCEEPQLRMKNNEAEEDYDDDLTDSMDVVR 360  
DB 320 MDLGOFLFCHISHOHDMGEAYVKVDSCEEPQLRMKNNEAEEDYDDDLTDSMDVVR 379  
QY 361 DDNSPSFIQIRSVAKKHPTWVHYIAAEEEDWDYAPLVLPADDRSYKSYLNNGPORIG 420  
DB 380 DDNSPSFIQIRSVAKKHPTWVHYIAAEEEDWDYAPLVLPADDRSYKSYLNNGPORIG 439  
QY 421 RYKVKVRFMAYTDFTFKTREAIOHESGILGPLLYGEVGDTLIIIFKNQASRPYNIYPHGI 480  
DB 440 RYKVKVRFMAYTDFTFKTREAIOHESGILGPLLYGEVGDTLIIIFKNQASRPYNIYPHGI 499  
QY 481 TDVRPLYSRLPKGVHLDKPPILPGEIFKYKMTVTVEDGPTKSDRCLTRYSSSFVNME 540  
DB 500 TDVRPLYSRLPKGVHLDKPPILPGEIFKYKMTVTVEDGPTKSDRCLTRYSSSFVNME 559  
QY 541 RDLASGLIGPLLICYKESVDQRGNQIMSDKRNVLFSVFDENRSWYLTENIQFLPNPAG 600  
DB 560 RDLASGLIGPLLICYKESVDQRGNQIMSDKRNVLFSVFDENRSWYLTENIQFLPNPAG 619  
QY 601 VQLEDPEFQASNMHMSINGVFDLSQLSVCLHEVAYWYIISIGAQTDFLSVFFSGYTFKH 660  
DB 620 VQLEDPEFQASNMHMSINGVFDLSQLSVCLHEVAYWYIISIGAQTDFLSVFFSGYTFKH 679  
QY 661 KMVYEDTLTLFPFSGETVFMSPENPGLWILGCHNSDFRNKGMTALLKVSSCDKNTGDYYE 720  
DB 680 KMVYEDTLTLFPFSGETVFMSPENPGLWILGCHNSDFRNKGMTALLKVSSCDKNTGDYYE 739

QY 721 DSYEDISAYLLSKNNAIEPRSFSONPPVLKRHOR-----EITRTTLOSDEBIDYDDTI 774  
DB 740 DSYEDISAYLLSKNNAIEPRSFSON-----SRHQVYRRGEITRTTLOSDEBIDYDDTI 795  
QY 775 SVMKKEDDIYDEDENQSPRSFQKTRHYFTAAVERLWDYGMSSSPHYLRNRAQSGSVP 834  
DB 796 SVMKKEDDIYDEDENQSPRSFQKTRHYFTAAVERLWDYGMSSSPHYLRNRAQSGSVP 855  
QY 835 QFKKVVQFBFTDGSFTQPLYRGELNEHLGLLGPYIRAEVEDNIMVTRNQASRPYFYSS 894  
DB 856 QFKKVVQFBFTDGSFTQPLYRGELNEHLGLLGPYIRAEVEDNIMVTRNQASRPYFYSS 915  
QY 895 LISYEDQOQABPRKNFKVNETKTYFKVQVHMAPTKDEFDCKAWAYFSDVLDKDVH 954  
DB 916 LISYEDQOQABPRKNFKVNETKTYFKVQVHMAPTKDEFDCKAWAYFSDVLDKDVH 975  
QY 955 SGLIGPLLYCHNTNLTNPAHGRQVTVQEFALFTTFIDETKSWYFTENWERNCRAPCNIOME 1014  
DB 976 SGLIGPLLYCHNTNLTNPAHGRQVTVQEFALFTTFIDETKSWYFTENWERNCRAPCNIOME 1035  
QY 1015 DPTFKENYRPHALINGYIMDTLPCLVMAQDQRIWYLLSMGNSNENIHSIHFSGHVFTVRKK 1074  
DB 1036 DPTFKENYRPHALINGYIMDTLPCLVMAQDQRIWYLLSMGNSNENIHSIHFSGHVFTVRKK 1095  
QY 1075 EYKMALYNLYPGVFTVEMLPKAGIWRVVECLIGELHAGMSTLFLVYSNKCQTPLGMA 1134  
DB 1096 EYKMALYNLYPGVFTVEMLPKAGIWRVVECLIGELHAGMSTLFLVYSNKCQTPLGMA 1155  
QY 1135 SGHIRDFQITASGOYQOWAPKLARLHYSINAWSTKEPFSWKVDLLAPMIHGIKTQG 1194  
DB 1156 SGHIRDFQITASGOYQOWAPKLARLHYSINAWSTKEPFSWKVDLLAPMIHGIKTQG 1215  
QY 1195 AROKFSLSYISQFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGINKNIFNPPIIARY 1254  
DB 1216 AROKFSLSYISQFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGINKNIFNPPIIARY 1275  
QY 1255 IRLHPTHYSIRSTLRMELMGCDLNSCMLPLGMSKASDAQITASSYFTNMFTWSPSKA 1314  
DB 1276 IRLHPTHYSIRSTLRMELMGCDLNSCMLPLGMSKASDAQITASSYFTNMFTWSPSKA 1335  
QY 1315 RLHLQGRSNARPOVNNPKWLOVDFOKTMKVTVGTQGVKSLTSMYKKEFLISSODG 1374  
DB 1336 RLHLQGRSNARPOVNNPKWLOVDFOKTMKVTVGTQGVKSLTSMYKKEFLISSODG 1395  
QY 1375 HQTWLFQFGKVKVFOGNQDSFTPVVNSLDPPLLTRYLRHPOSWVHQAIRMEVLGCEA 1434  
DB 1396 HQTWLFQFGKVKVFOGNQDSFTPVVNSLDPPLLTRYLRHPOSWVHQAIRMEVLGCEA 1455  
QY 1435 QDLY 1438  
DB 1456 QDLY 1459  
  
RESULT 10  
AAE10827  
ID AAE10827 standard; Protein; 1459 AA.  
XX  
XX AAE10827;  
XX AC  
XX 18-DEC-2001 (first entry)  
XX  
XX Human factor VIII mutin encoded by vector pTGF8-1.  
XX  
XX Human; haemostatic; coagulant; blood clotting factor; factor VIII;  
XX factor IX; therapy; haemophilia A; mutant; mutin.  
XX  
XX Homo sapiens.  
XX Synthetic.  
XX  
XX Key Location/Qualifiers  
XX 1..19 Peptide  
XX /label= signal\_peptide  
XX 20..1459 Protein  
XX FT



FT Misc-difference 181 /product= "Human mature factor VIII mutuin"  
FT /note= "Wild type Val substituted with Ala"  
FT Misc-difference 1138 /note= "Wild type Val substituted with Asn"  
FT /note= "Wild type Ser substituted with Asn"  
FT Misc-difference 1350 /note= "Wild type Val substituted with Glu"  
FT  
XX  
FN WO200170968-A2.  
XX  
PD 27-SEP-2001.  
XX  
XX 21-MAR-2001; 2001WO-EP032220.  
XX  
XX 22-MAR-2000; 2000EP-0106225.  
PR 08-MAY-2000; 2000US-203249P.  
XX  
XX (OCTA-) OCTAGENE GMBH.  
XX  
XX Hauser C, Hoerster A, Schroeder C, Lehnerer M;  
XX  
XX WPI; 2001-590175/66.  
DR N-PSDB; AAD18172.  
XX  
XX Recombinantly producing human blood coagulation factors VIII and IX for  
PT use in treating hemophilia -  
PT  
XX  
XX Claim 17; Page 69-73; 104pp; English.  
XX  
XX The present invention relates to an improved method for the production  
CC of recombinant human blood clotting factors, especially factors VIII and  
CC IX, using an immortalised cell line stably expressing viral transcription  
CC activators and carrying a vector comprising a promoter and a sequence  
CC encoding the blood coagulation factor. The factor VIII mutuin or a gene  
CC transfer vector is used in the preparation of agents for treating  
CC haemophilia, especially haemophilia A. The present sequence is  
CC human factor VIII mutuin encoded by vector pTGF8-1.  
XX  
XX Sequence 1459 AA;  
SQ  
Query Match 99.0%; Score 7613; DB 22; Length 1459;  
Best Local Similarity 98.9%; Pred. No. 0;  
Matches 1428; Conservative 1; Mismatches 5; Indels 10; Gaps 2;  
QY 1 ATRRYLGAVELSDYMQSDLGELPVDARFPVRPKSPFNNTSVYKKTLPVEFTVHLFN 60  
DB 20 ATRRYLGAVELSDYMQSDLGELPVDARFPVRPKSPFNNTSVYKKTLPVEFTDHLFN 79  
QY 61 IAKRPPPMGLGPTIOAEVYDVTVTITKNASHPVSLHAGVSVYKASEGAEYDDQTSQ 120  
DB 80 IAKRPPPMGLGPTIOAEVYDVTVTITKNASHPVSLHAGVSVYKASEGAEYDDQTSQ 139  
QY 121 REKEDKVPFGGSHYVWQVLKENGPMASDPLCLTYSLYSHVDLVKOLNSGLIGALLVCR 180  
DB 140 REKEDKVPFGGSHYVWQVLKENGPMASDPLCLTYSLYSHVDLVKOLNSGLIGALLVCR 199  
QY 181 EGSIAKEXTQTLHKFILLFAVEDEKSWHSETKNSLMQDRDAASAPWPKMHTVNGYNNR 240  
DB 200 EGSIAKEXTQTLHKFILLFAVEDEKSWHSETKNSLMQDRDAASAPWPKMHTVNGYNNR 259  
QY 241 SLPLGLIGCHRKSYYWHVYGMGTPPVHVSIFLEGHTFLVRNHRQASLEISPTIFLTAQTLL 300  
DB 260 SLPLGLIGCHRKSYYWHVYGMGTPPVHVSIFLEGHTFLVRNHRQASLEISPTIFLTAQTLL 319  
QY 301 MDLGOFLFCHTSSHHQDGEAYVKVDSCEPPEQLRMKNNEAEYDDDLTDSEMDVVRP 360  
DB 320 MDLGOFLFCHTSSHHQDGEAYVKVDSCEPPEQLRMKNNEAEYDDDLTDSEMDVVRP 379  
QY 361 DDNSPFSFIQIRSVAKGHPKTVVHYIAAEEEDWDYAPLVAPDDRYSYKSQYLNNGPQIRG 420  
DB 380 DDNSPFSFIQIRSVAKGHPKTVVHYIAAEEEDWDYAPLVAPDDRYSYKSQYLNNGPQIRG 439  
QY 421 RYKKVRPMAYTDETFKTRAIQHESGILGPLLYGEGVDTLLIIIFKQASRPYNIYPHGI 480

DB 440 RYKKVRPMAYTDETFKTRAIQHESGILGPLLYGEGVDTLLIIIFKQASRPYNIYPHGI 499  
QY 481 TDVRPLYSRRLPKGVKHLKDFPILPGBIFKYKMTVTVEDGPTKSDPCLTRYSSFFVME 540  
DB 500 TDVRPLYSRRLPKGVKHLKDFPILPGBIFKYKMTVTVEDGPTKSDPCLTRYSSFFVME 559  
QY 541 RDLASGLIGPLLCYKESVDQGNQINMSDKRNVILFSVFDENRSWYLTENIQRLPNPAG 600  
DB 560 RDLASGLIGPLLCYKESVDQGNQINMSDKRNVILFSVFDENRSWYLTENIQRLPNPAG 619  
QY 601 VOLEDEPFOASNIHMSINGVYFDSLOLSVCLHEVAYWYILSIGAQTDPLSVFFSGYTFKH 660  
DB 620 VOLEDEPFOASNIHMSINGVYFDSLOLSVCLHEVAYWYILSIGAQTDPLSVFFSGYTFKH 679  
QY 661 KMVYEDTLTLFPFSGETVFMSENPGWLILGCHNSDFNRGMTALLKVSSCDKNTGDYYE 720  
DB 680 KMVYEDTLTLFPFSGETVFMSENPGWLILGCHNSDFNRGMTALLKVSSCDKNTGDYYE 739  
QY 721 DSYEDISAYLLSKNNAIEPRSFQNPVLPKHQR-----BITRTTLOSQEBEIDYDDTI 774  
DB 740 DSYEDISAYLLSKNNAIEPRSFQNPVLPKHQR-----SRHQAYRYRGEITRTTLOSQEBEIDYDDTI 795  
QY 775 SVMKKEDFDIYDEDENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNRAQSQSV 834  
DB 796 SVMKKEDFDIYDEDENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNRAQSQSV 855  
QY 835 QFKKVPFQEFDTGSGFTQPLYRGELNEHLGLGPYIRAEVDNIIMVTFRNQASRPYSFSS 894  
DB 856 QFKKVPFQEFDTGSGFTQPLYRGELNEHLGLGPYIRAEVDNIIMVTFRNQASRPYSFSS 915  
QY 895 LISYEEORQGAERPKNFVKPNETKTYFWKVQHHMAPTKDFCKAWAYSDVDLEKDVH 954  
DB 916 LISYEEORQGAERPKNFVKPNETKTYFWKVQHHMAPTKDFCKAWAYSDVDLEKDVH 975  
QY 955 SGLIGPLLVCHTNTLNPAHGRQVTVQEFALFFTIFDETCKSWYFTENMERNCRAPCNIQME 1014  
DB 976 SGLIGPLLVCHTNTLNPAHGRQVTVQEFALFFTIFDETCKSWYFTENMERNCRAPCNIQME 1035  
QY 1015 DPTFKENYRFAHNGYIMDTLPLGLVMAQDQIRWYLLSMGNSNENIHSIHSGHVFTVRKK 1074  
DB 1036 DPTFKENYRFAHNGYIMDTLPLGLVMAQDQIRWYLLSMGNSNENIHSIHSGHVFTVRKK 1095  
QY 1075 BEYKQALYNLYPGVFETVEMLPKAGIWRVECLIGELHAGMSTLFLVYSNKCQTPLGMA 1134  
DB 1096 BEYKQALYNLYPGVFETVEMLPKAGIWRVECLIGELHAGMSTLFLVYSNKCQTPLGMA 1155  
QY 1135 SGHIRDFQITASGOYGOWAPKILARLHYSGSINAMSTKEPPSWIKVDLLAPMIHGIKTQG 1194  
DB 1156 SGHIRDFQITASGOYGOWAPKILARLHYSGSINAMSTKEPPSWIKVDLLAPMIHGIKTQG 1215  
QY 1195 ARQKFSSLYISQFIIMYSLDGKKQWYRGNSTGTLMVFPFNVDSGINKHNIFFNPPIIARY 1254  
DB 1216 ARQKFSSLYISQFIIMYSLDGKKQWYRGNSTGTLMVFPFNVDSGINKHNIFFNPPIIARY 1275  
QY 1255 IRLHPTHYSIRSTRMELMGCDLNSCMLPGMESKASDAQITASSYFTNMFATWSPSKA 1314  
DB 1276 IRLHPTHYSIRSTRMELMGCDLNSCMLPGMESKASDAQITASSYFTNMFATWSPSKA 1335  
QY 1315 RLHIQGRSNARWPOVNNPKWLQYDFOKTMKVGTQGVKSLTSMYVKEFLISSSDQG 1374  
DB 1336 RLHIQGRSNARWPOVNNPKWLQYDFOKTMKVGTQGVKSLTSMYVKEFLISSSDQG 1395  
QY 1375 HQWTLFFQNGKVKVFGQNDQSFTEPVNSLOPPLLTLYLRHQPQSWHQAIALRMEVLGCEA 1434  
DB 1396 HQWTLFFQNGKVKVFGQNDQSFTEPVNSLOPPLLTLYLRHQPQSWHQAIALRMEVLGCEA 1455  
QY 1435 QDLY 1438  
DB 1456 QDLY 1459  
RESULT 11

AAP80265  
ID AAP80265 standard; protein; 1516 AA.

AC AAP80265;

XX 25-MAR-2003 (updated)

DT 10-OCT-1990 (first entry)

XX Modified factor VIII:C sequence with the Q744-D1563 deletion.

XX Modified factor VIII:C; maturation polypeptide; haemophilia;

KW blood coagulation; QD deletion.

XX Homo sapiens.

XX WO8800831-A.

XX 11-FEB-1988.

PF 31-JUL-1987; 87WO-US01814.

XX 01-AUG-1986; 86US-0893375.

XX (BIOJ ) BIOGEN NV.

PA (PASE/) PASEK M P.

XX Pasek MP;

XX WPI; 1988-049866/07.

DR N-PSDB; AAN80444.

XX New DNA sequences encoding modified factor VIII:C - with deletion of DNA  
PT encoding maturation polypeptide, useful for high yield transformation.

XX Claim 3; Page 51-52-53-54; 97pp; English.

XX A major part of the sequence encoding the maturation polypeptide of

CC factor VIII:C is deleted, i.e. Gln 744 - Asp 1563. The QD deletion

CC retains approximately 90 amino acids of the maturation polypeptide

CC (four amino acids at the N-terminal end and 86 amino acids at

CC the C-terminal end). The full length Factor VIII:C cDNA has two

CC changes with respect to the published sequence (EPO application 160457):

CC Crg to Crg at Leu 242 and TTC to CTC change at amino acid residue 1880

CC (Phe to Leu).

CC The product is produced in approx. 20 times higher

CC yields than previous recombinant produced factor VIII:C and are more

CC easily purified. The peptide is used for treating haemophilia A, both

CC acute and prolonged bleeding.

CC See also AAN80446 and AAN80447.

CC (Updated on 25-MAR-2003 to correct PA field.)

XX Sequence 1516 AA;

SQ

Query Match 99.08; Score 7611.5; DB 9; Length 1516;

Best Local Similarity 94.74; Pred. No. 0;

Matches 1434; Conservative 1; Mismatches 3; Indels 77; Gaps 2;

QY 1 ATRRYLGAVELSDVMQSDGLGELPVDARPPRPVPSFPNTSVYKTLFVETVHLFN 60

DB 2 ATRRYLGAVELSDVMQSDGLGELPVDARPPRPVPSFPNTSVYKTLFVETVHLFN 61

QY 61 IAKPRPPWMLGPTTIAQEVYDVTVTITLKNASHPVSILHVGVSVMKASGAEYDDQTSQ 120

DB 62 IAKPRPPWMLGPTTIAQEVYDVTVTITLKNASHPVSILHVGVSVMKASGAEYDDQTSQ 121

QY 121 REKEDDKVFGSGSYTYVQVLKENGPMASDPLCLTYSYLHVDLVKDLNSGLIGALLVCR 180

DB 122 REKEDDKVFGSGSYTYVQVLKENGPMASDPLCLTYSYLHVDLVKDLNSGLIGALLVCR 181

QY 181 EGSIAKSKTQTLHKFILLFAVFDGSKSWHSETKNSLMQDRDAASARAWPKMHTVGVYNR 240

DB 182 EGSIAKSKTQTLHKFILLFAVFDGSKSWHSETKNSLMQDRDAASARAWPKMHTVGVYNR 241

QY 241 S-LPLGLIGCHRSVYWHVIGMGTTPVHSIFLEGHTFLVRNHRQASLEISPTFLTAQTL 299

DB 242 SLLPGLIGCHRSVYWHVIGMGTTPVHSIFLEGHTFLVRNHRQASLEISPTFLTAQTL 301

QY 300 LMDLGQFLFCHISSHQHDMAYVKVDSCEPQPOLRMKNNEARAYDDDLTDTSEMDVVR 359

DB 302 LMDLGQFLFCHISSHQHDMAYVKVDSCEPQPOLRMKNNEARAYDDDLTDTSEMDLVR 361

QY 360 FDDNSPFIQIRSVAKGHPKTVWHVIAAEEEDWDYAPLVLAPDDRSYKSYQLNNGPQRI 419

DB 362 FDDNSPFIQIRSVAKGHPKTVWHVIAAEEEDWDYAPLVLAPDDRSYKSYQLNNGPQRI 421

QY 420 GRKYKVRMAYTDETFKTRAIQHESGILGPLLYGEGVDTLLIIIPKQASRYNIYPHG 479

DB 422 GRKYKVRMAYTDETFKTRAIQHESGILGPLLYGEGVDTLLIIIPKQASRYNIYPHG 481

QY 480 ITDVRPLYSRRLPKGVKHLKDPFIPGEIFKYKWTVTVEDGPTKSDPRCLTRYYSFVNM 539

DB 482 ITDVRPLYSRRLPKGVKHLKDPFIPGEIFKYKWTVTVEDGPTKSDPRCLTRYYSFVNM 541

QY 540 ERDLASGLIGPLLI CYKESVDQGNQIMSDKRNVLFSVFDENRSWYLTENIQRFNP 599

DB 542 ERDLASGLIGPLLI CYKESVDQGNQIMSDKRNVLFSVFDENRSWYLTENIQRFNP 601

QY 600 GVQLEDPEFQASNIMHSINGYVFDLSQLSVCLHEVAYWYLSIGAQTFDLSVFSGYTFK 659

DB 602 GVQLEDPEFQASNIMHSINGYVFDLSQLSVCLHEVAYWYLSIGAQTFDLSVFSGYTFK 661

QY 660 HKMVEDTTLTLPFSGETVFMSENPLGTLGCHNSDFRNRMGTALLKVVSSCDKNTGDYY 719

DB 662 HKMVEDTTLTLPFSGETVFMSENPLGTLGCHNSDFRNRMGTALLKVVSSCDKNTGDYY 721

QY 720 EDSYEDISAYLLSKNNAIBRSPF----- 742

DB 722 EDSYEDISAYLLSKNNAIBRSPSQDPLAWNHYGTQIPKEWKSQEKSEKTAFFKKDT 781

QY 743 -----SONPVVLRKHOREITRITLQ 763

DB 782 ILSINACESHATAINEGQNKPEIEVWAKQGRTERLCSQNPVPLKRHOREITRITLQ 841

QY 764 DQBEIDYDDTISVEMKEDFDIYDEBENOSPRSQKTRHYFTAAVERLDYGMSSSPHV 823

DB 842 DQBEIDYDDTISVEMKEDFDIYDEBENOSPRSQKTRHYFTAAVERLDYGMSSSPHV 901

QY 824 LRNRAQSGSVQPKVVFQEFDTGFTQPLRYGELNEHLGLGPIYAEVEDNIMVTFRN 883

DB 902 LRNRAQSGSVQPKVVFQEFDTGFTQPLRYGELNEHLGLGPIYAEVEDNIMVTFRN 961

QY 884 QASRPYSFYSLLSYBEDQOQAEPRKFNFKVNETKTYFMKVQHMAPTKDEPDCCKAWAY 943

DB 962 QASRPYSFYSLLSYBEDQOQAEPRKFNFKVNETKTYFMKVQHMAPTKDEPDCCKAWAY 1021

QY 944 FSDVDLEKDVHSLGILGLLVCHTNTLNPAHGRVTVQEPALPFTIPEKTSWTFENMER 1003

DB 1022 FSDVDLEKDVHSLGILGLLVCHTNTLNPAHGRVTVQEPALPFTIPEKTSWTFENMER 1081

QY 1004 NCRAPCNIOEMDPTFKENYFHAINGYIMDTLPLGVMAQDQIRWYLLSGNSNETHSIH 1063

DB 1082 NCRAPCNIOEMDPTFKENYFHAINGYIMDTLPLGVMAQDQIRWYLLSGNSNETHSIH 1141

QY 1064 FSGHVTFRKKEBYKVALYNLYPGVPEVEMLPKAGIWRVECLIGELHAGMSTFLVY 1123

DB 1142 FSGHVTFRKKEBYKVALYNLYPGVPEVEMLPKAGIWRVECLIGELHAGMSTFLVY 1201

QY 1124 SNKCQPLGNASGHIEDFQITASGOYQWAPKLARLHYSGSINAWSTKEPFSIKVDLLA 1183

DB 1202 SNKCQPLGNASGHIEDFQITASGOYQWAPKLARLHYSGSINAWSTKEPFSIKVDLLA 1261

QY 1184 PMIHGKITQCARQKFSLLYSIQFIWYSLDGKKWOTYRGNSTGLMVFNGVDSGSIKH 1243

DB 1262 PMIHGKITQCARQKFSLLYSIQFIWYSLDGKKWOTYRGNSTGLMVFNGVDSGSIKH 1321

QY 1244 NIFNPPIIARYIRLHPHTHYSTIRSTRMLMGLMGLNSCSMPLGMESKAISDAQITASSYFT 1303

|||||  
Db 1322 NIFNPPIIARIIRLHTYSIRSTURMELMGCDLNSCSMPLEMSKALSDAQITASSYFT 1381  
Qy 1304 NMFATWSPSKARLHLQGRSNAWRPQVNNPKFQWLDVDFQRTMKVGTGVTGQVKSLTSMYV 1363  
Db 1382 NMFATWSPSKARLHLQGRSNAWRPQVNNPKFQWLDVDFQRTMKVGTGVTGQVKSLTSMYV 1441  
Qy 1364 KEFLISSQDGHQWTLFFQNGKVKVFOGQNDSTFVNSLDPPLLTRYLRHHPQSWVHQI 1423  
Db 1442 KEFLISSQDGHQWTLFFQNGKVKVFOGQNDSTFVNSLDPPLLTRYLRHHPQSWVHQI 1501  
Qy 1424 ALRMEVLGCEAODLY 1438  
Db 1502 ALRMEVLGCEAODLY 1516  
RESULT 12  
ID AAB48842  
ID AAB48842 standard; protein; 1424 AA.  
XX AAB48842;  
AC AAB48842;  
XX 13-MAR-2001 (first entry)  
XX Mutant mature human factor VIII, SEQ ID NO:5.  
XX Factor VIII; human; B domain; LRP-mediated plasma clearance;  
KW receptor-dependent clearance; receptor-independent clearance;  
KW half-life; haemophilia; mutant; mutein.  
XX Homo sapiens.  
XX WO200071714-A2.  
FN 30-NOV-2000.  
PD 24-MAY-2000; 2000WO-US14111.  
PF 24-MAY-1999; 99US-0135847.  
PR 24-MAY-1999; 99US-0135847.  
XX (AMNA-) AMERICAN NAT RED CROSS.  
PA Saenko EL, Strickland DK;  
PI WPI; 2001-025163/03.  
XX Factor VIII mutants having increased half-life useful for treating  
PT hemophilia, comprise one or more amino acid substitutions in the A2  
PT and/or C2 domain of factor VIII -  
XX Claim 9; Fig 2A-B; 121pp; English.  
XX The invention relates to human factor VIII mutants comprising an amino  
CC acid substitution at one or more positions in the A2 domain and/or an  
CC amino acid substitution at one or more positions in the C2 domain.  
CC The invention also encompasses a factor VIII mutant which lacks a B  
CC domain (AAB48842). The factor VIII mutants have an increased half-life  
CC in the bloodstream. The A2 domain mutants exhibit reduced LRP-dependent  
CC (receptor-dependent) clearance of factor VIII, while C2 domain mutants  
CC have reduced receptor-independent clearance. The invention also relates  
CC to a method of using RAP (receptor associated protein), a protein which  
CC inhibits LRP (low density lipoprotein related protein)-mediated ligand  
CC internalisation, to increase the half-life of factor VIII. The mutant  
CC factor VIII proteins, and nucleotides encoding them, are useful  
CC for treating haemophilia. RAP, LRP-binding RAP mutants or fragments, and  
CC nucleic acids encoding them may also be used in the treatment of  
CC haemophilia, in combination with a mutant factor VIII protein or DNA of  
CC the invention. The invention provides means of increasing the half-life  
CC of factor VIII by reducing its clearance from plasma. The present  
CC sequence represents a mutant mature human factor VIII which lacks a B  
CC domain.  
XX Sequence 1424 AA;  
SQ

Query Match 98.7%; Score 7592; DB 22; Length 1424;  
Best Local Similarity 99.0%; Pred. No. 0;  
Matches 1423; Conservative 0; Mismatches 1; Indels 14; Gaps 1;  
Qy 1 ATRYYILGAVELSDWYQSDLGELPVDARPPRPVKSPFPNTSVVYKKTLEVEFTVHLFN 60  
Db 1 ATRYYILGAVELSDWYQSDLGELPVDARPPRPVKSPFPNTSVVYKKTLEVEFTVHLFN 60  
Qy 61 IAKPRPPMGLLGTPTIAQEVYDVTVITLKNASHPVSLHVGVSYKASEGAEYDDQTSQ 120  
Db 61 IAKPRPPMGLLGTPTIAQEVYDVTVITLKNASHPVSLHVGVSYKASEGAEYDDQTSQ 120  
Qy 121 REKEDDKVPFGSHTYVQVLKENGPMASDPLCTYSYLSHVDLVKOLNSGLIGALLVCR 180  
Db 121 REKEDDKVPFGSHTYVQVLKENGPMASDPLCTYSYLSHVDLVKOLNSGLIGALLVCR 180  
Qy 181 EGSIAKETQTLHRKFIILLFAVFDGKSWHSSTKNSLMQDRDAASARAWPKMHTVGVYNR 240  
Db 181 EGSIAKETQTLHRKFIILLFAVFDGKSWHSSTKNSLMQDRDAASARAWPKMHTVGVYNR 240  
Qy 241 SLPLGLIGCHRSVYVHVGITTPPEVHSIFLEGHTFLVRNHRQASLEISPIITFLTAQTLL 300  
Db 241 SLPLGLIGCHRSVYVHVGITTPPEVHSIFLEGHTFLVRNHRQASLEISPIITFLTAQTLL 300  
Qy 301 MDLQQLFCHISSHQHDMGMEAYVKVDSCEPQOLRMKNNEEADYDDDLTSEMDVVRF 360  
Db 301 MDLQQLFCHISSHQHDMGMEAYVKVDSCEPQOLRMKNNEEADYDDDLTSEMDVVRF 360  
Qy 361 DDNSPSFIQIRSVAKKHPKTHVHIAAEEDWDYAPLVLAPDRSKYSQVLYNNGPQIG 420  
Db 361 DDNSPSFIQIRSVAKKHPKTHVHIAAEEDWDYAPLVLAPDRSKYSQVLYNNGPQIG 420  
Qy 421 RYKVKRFMAYTDETFKTREAIQESGILGLPGLLYGEVGDITLLIIFKQASPYNIYPHGI 480  
Db 421 RYKVKRFMAYTDETFKTREAIQESGILGLPGLLYGEVGDITLLIIFKQASPYNIYPHGI 480  
Qy 481 TDVRPLYSRRLPKGVKHLKOPPIIPGEIFKTKWTVTVEGDKTSDPRCLTRYYSFVNNE 540  
Db 481 TDVRPLYSRRLPKGVKHLKOPPIIPGEIFKTKWTVTVEGDKTSDPRCLTRYYSFVNNE 540  
Qy 541 RDLASGLIGPLLIICYKESVDQGNQIMSDKENVILFSVFDENRSMYLTENORLEPNAG 600  
Db 541 RDLASGLIGPLLIICYKESVDQGNQIMSDKENVILFSVFDENRSMYLTENORLEPNAG 600  
Qy 601 VQLEDPEFQASNIMHSINGYVFDLSQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKH 660  
Db 601 VQLEDPEFQASNIMHSINGYVFDLSQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKH 660  
Qy 661 KMVYEDTLTLPFSGETVFMSENPGWLILGCHNSDFNRGWTALLKVSSCDKNTGDYYE 720  
Db 661 KMVYEDTLTLPFSGETVFMSENPGWLILGCHNSDFNRGWTALLKVSSCDKNTGDYYE 720  
Qy 721 DSYEDIISAYLLSKNATEPRSFQNPVVLKHHQREITRTTLQSDQERIDYDDTISVEMKK 780  
Db 721 DSYEDIISAYLLSKNATEPRSFQNPVVLKHHQREITRTTLQSDQERIDYDDTISVEMKK 780  
Qy 781 EDFDIYDEDNQSPRSFQKTRHYFIAAVERLWDYGMSSSPHYLRNRAQSGSVQFKKVV 840  
Db 781 EDFDIYDEDNQSPRSFQKTRHYFIAAVERLWDYGMSSSPHYLRNRAQSGSVQFKKVV 840  
Qy 841 FQETDGSFTQPLRYGELNEHGLLGPYIRAEVEDNIMVTFRNQASRPYSYSSLSIYEE 900  
Db 841 FQETDGSFTQPLRYGELNEHGLLGPYIRAEVEDNIMVTFRNQASRPYSYSSLSIYEE 900  
Qy 901 DOROGAPRKNFVKPNETKTYFMVQHHMATKDFCDKAWAYSDVDLEKDVHSLIGP 960  
Db 901 DOROGAPRKNFVKPNETKTYFMVQHHMATKDFCDKAWAYSDVDLEKDVHSLIGP 960  
Qy 961 LLVCHTNTLNPAHGRQVTVQEFALFFITFDFTKSWYFTENMERNCRAPCNIQMEDPPTKE 1020  
Db 961 LLVCHTNTLNPAHGRQVTVQEFALFFITFDFTKSWYFTENMERNCRAPCNIQMEDPPTKE 1020  
Qy 947 LLVCHTNTLNPAHGRQVTVQEFALFFITFDFTKSWYFTENMERNCRAPCNIQMEDPPTKE 1006  
Db 947 LLVCHTNTLNPAHGRQVTVQEFALFFITFDFTKSWYFTENMERNCRAPCNIQMEDPPTKE 1006

QY 1021 NYRPHAINGYIMDTLPGLVMAQDQIRIRWYLLSMGSENENIHSIHPSGHVFTVRKKEBYKMA 1080  
DB 1007 NYRPHAINGYIMDTLPGLVMAQDQIRIRWYLLSMGSENENIHSIHPSGHVFTVRKKEBYKMA 1066  
QY 1081 LYNLYPGVFETVEMLPKAGIWRVECLIGELHAGMSTFLVYSNKQOTPLGMAASHIRD 1140  
DB 1067 LYNLYPGVFETVEMLPKAGIWRVECLIGELHAGMSTFLVYSNKQOTPLGMAASHIRD 1126  
QY 1141 FQITASQYQGWAPKPLARLHYSGSINAWSTKEPFSWIKVDLAPMIITHGIKTQCARQKFS 1200  
DB 1127 FQITASQYQGWAPKPLARLHYSGSINAWSTKEPFSWIKVDLAPMIITHGIKTQCARQKFS 1186  
QY 1201 SLTISQFIIMYSLDGKQWYRGNTGTLMVFFGNVDSSGKKNIFNPPIIARIYRLHPT 1260  
DB 1187 SLTISQFIIMYSLDGKQWYRGNTGTLMVFFGNVDSSGKKNIFNPPIIARIYRLHPT 1246  
QY 1261 HYSIRSTRMELMCDLNSCMPLGMSKALSDAQITASSYFTNMFATWSPSKARLHLOQ 1320  
DB 1247 HYSIRSTRMELMCDLNSCMPLGMSKALSDAQITASSYFTNMFATWSPSKARLHLOQ 1306  
QY 1321 RSNARWQVNNPKWLOVDFOKTMKVGTGVTGQVKSLLTSMYVKEFLISSQDGHQWTLF 1380  
DB 1307 RSNARWQVNNPKWLOVDFOKTMKVGTGVTGQVKSLLTSMYVKEFLISSQDGHQWTLF 1366  
QY 1381 FONGKVKVQGNQDSFTPVVNSLDPPLLTRYLRTHPSQSWHQIALRMEVLGCEAQDLY 1438  
DB 1367 FONGKVKVQGNQDSFTPVVNSLDPPLLTRYLRTHPSQSWHQIALRMEVLGCEAQDLY 1424

RESULT 13  
AAO18622  
ID AAO18622 standard; Protein; 1424 AA.  
AC AAO18622;  
XX 24-OCT-2002 (first entry)  
DT Human mature B-domainless factor VIII.  
DE Human; factor VIII; fVIII; half-life; mutant; haemophilia;  
XX heparan sulfate proteoglycan-mediated clearance; RAP;  
KW receptor-associated protein; haemostatic; gene therapy;  
KW alpha2 macroglobulin receptor-associated protein.  
XX Homo sapiens.  
OS Key Location/Qualifiers  
PH Domain 373..740  
FT /label= A2 domain  
FT Region 484..509  
FT /label= LRP\_binding\_region  
XX WO200260951-A2.  
XX 08-AUG-2002.  
XX 11-JAN-2002; 2002WO-US00583.  
XX 12-JAN-2001; 2001US-260904P.  
XX (AMNA-) AMERICAN NAT RED CROSS.  
FA Saenko EL, Sarafanov AG;  
XX WPI; 2002-608501/65.  
XX New mutant factor VIII with reduced sulfate proteoglycan  
PT (HSPG)-dependent or receptor-independent clearance and procoagulant  
PT activity for treating hemophilia -  
XX Claim 9; Fig 12; 161pp; English.  
XX The present invention relates to a mutant factor VIII protein with

CC reduced sulfate proteoglycan (HSPG)-dependent or receptor-independent  
CC clearance and procoagulant activity, which has a nonconservative amino  
CC acid substitution at one or more positions in the A2 domain consisting of  
CC Lys(380, 512, 556, 570 or 659) or Arg(490, 527, 562 or 571) or the C2  
CC domain relative to the wild-type. The mutant factor VIII or the  
CC polynucleotide encoding it and a receptor-associated protein (alpha2  
CC macroglobulin receptor-associated protein or RAP) are useful for treating  
CC haemophilia. The mutated protein has a longer half-life. The present  
CC sequence is the human mature B-domainless factor VIII protein.  
XX  
SQ Sequence 1424 AA;  
Query Match 98.7%; Score 7592; DB 23; Length 1424;  
Best Local Similarity 99.0%; Pred. No. 0;  
Matches 1423; Conservative 0; Mismatches 1; Indels 14; Gaps 1;  
QY 1 ATRRYILGAVELSDWYMQSDLGELPVDARPPVPKSPFNTSVYVKTILFVFTVHLFN 60  
DB 1 ATRRYILGAVELSDWYMQSDLGELPVDARPPVPKSPFNTSVYVKTILFVFTVHLFN 60  
QY 61 IAKPRPPMGLGPTIOAEVYDVTVTILKNMASHPVSLHAGVSYWKASBGAEYDDQTSQ 120  
DB 61 IAKPRPPMGLGPTIOAEVYDVTVTILKNMASHPVSLHAGVSYWKASBGAEYDDQTSQ 120  
QY 121 REKEDDKVFGSGHTYVQVLKENGPMASDPLCLITYSLSHVDLVKDLNGLIGALLVCR 180  
DB 121 REKEDDKVFGSGHTYVQVLKENGPMASDPLCLITYSLSHVDLVKDLNGLIGALLVCR 180  
QY 181 EGSILAKEKTQTLHKFILLFAVDFEGKSWHSETKNSLMQDRDAASARAWPKMHTVGVNVR 240  
DB 181 EGSILAKEKTQTLHKFILLFAVDFEGKSWHSETKNSLMQDRDAASARAWPKMHTVGVNVR 240  
QY 241 SLPLIGICHRKSVYVHVGMTTPEVHSIFLEGHTFLVRNHRQASLEISPIITFLTAQLL 300  
DB 241 SLPLIGICHRKSVYVHVGMTTPEVHSIFLEGHTFLVRNHRQASLEISPIITFLTAQLL 300  
QY 301 MDLGOFLFCHISHQHDGMEAYVKVDSCEEPQLRMKNBEAEYDDDLTDSMDVVRP 360  
DB 301 MDLGOFLFCHISHQHDGMEAYVKVDSCEEPQLRMKNBEAEYDDDLTDSMDVVRP 360  
QY 361 DDNSPSFIOIRSVAKKHPTWVHYIAAEEDMDYAPLVLPDDRYSKSOVLNNGPORIG 420  
DB 361 DDNSPSFIOIRSVAKKHPTWVHYIAAEEDMDYAPLVLPDDRYSKSOVLNNGPORIG 420  
QY 421 RYKVKVRFMAYTDETFKTREAIQHSIGILGPLLYGEVDTLLIIFKNQASRPYNIYPHGI 480  
DB 421 RYKVKVRFMAYTDETFKTREAIQHSIGILGPLLYGEVDTLLIIFKNQASRPYNIYPHGI 480  
QY 481 TDVRPLYSRRLPKGVKHLKDPPIIPGEIFKYKWTVTVEDEGPTKSDPCLTRYISGFVME 540  
DB 481 TDVRPLYSRRLPKGVKHLKDPPIIPGEIFKYKWTVTVEDEGPTKSDPCLTRYISGFVME 540  
QY 541 RDLASGLIGPLLCYKESVDQGNQIMSDKRNVLFSVFDENRSWYLTENIQRLFPAPAG 600  
DB 541 RDLASGLIGPLLCYKESVDQGNQIMSDKRNVLFSVFDENRSWYLTENIQRLFPAPAG 600  
QY 601 VQLEDPEFQASNMHSINGVYFDSLQSVCLHEVAYWYIISIGAQTDFLSVFFSGYTFKH 660  
DB 601 VQLEDPEFQASNMHSINGVYFDSLQSVCLHEVAYWYIISIGAQTDFLSVFFSGYTFKH 660  
QY 661 KMVYEDTLTLFPFSGETVFMSENPGWLWILGCHNSDFRNKGMTALLKVSSCDKNTGYE 720  
DB 661 KMVYEDTLTLFPFSGETVFMSENPGWLWILGCHNSDFRNKGMTALLKVSSCDKNTGYE 720  
QY 721 DSVYEDISAYLLSKNNAIEPRSFQNPVPLXKHOREITRTTLQSDQOEIDYDDTISVEMKK 780  
DB 721 DSVYEDISAYLLSKNNAIEP-----REITRTTLQSDQOEIDYDDTISVEMKK 766  
QY 781 EDFDIYDEDENQSPRSFQKTRHYFIAAVERLMDYGMSSSPHYLRNQAQSGSVQFQKVV 840  
DB 767 EDFDIYDEDENQSPRSFQKTRHYFIAAVERLMDYGMSSSPHYLRNQAQSGSVQFQKVV 826  
QY 841 FQEFDTGDSFTQPLYRGELNBEHLGLLPYIRAEVEDNIMVTFRNQASRPYFSYSLISYEE 900

Db 827 FQFTDGSFTQPLRGELNEHLGLLPYIRAEVEDNIMVTFRNOASRPYSFSSLSYEE 886  
QY 901 DORQGAEPKPNFVKNETTYFKWQHMAPKDEPDCKAWAFSDVDLEKDVHSLGIP 960  
Db 887 DORQGAEPKPNFVKNETTYFKWQHMAPKDEPDCKAWAFSDVDLEKDVHSLGIP 946  
QY 961 LLVCHTNTLNPAHGRQVTVQEFALPFTTIFDETKSWYFTENMERNCRAPCNIQMEDPTFK 1020  
Db 947 LLVCHTNTLNPAHGRQVTVQEFALPFTTIFDETKSWYFTENMERNCRAPCNIQMEDPTFK 1006  
QY 1021 NYRFAHNGIYIMDTLPCLYMAQDORIRWYLLSGNSNENIHSIFSGHVTVRKCEYKMA 1080  
Db 1007 NYRFAHNGIYIMDTLPCLYMAQDORIRWYLLSGNSNENIHSIFSGHVTVRKCEYKMA 1066  
QY 1081 LYNLYPGVETVEMLPSPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQPLGMAASHIRD 1140  
Db 1067 LYNLYPGVETVEMLPSPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQPLGMAASHIRD 1126  
QY 1141 FOITASGOYQWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMIHGIKTQGARQKFS 1200  
Db 1127 FOITASGOYQWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMIHGIKTQGARQKFS 1186  
QY 1201 SYIISOPIIIMYSLDGKKWQTYRGNSTGTLWVFGNVDSGIGKHINFPPIIARYIRLHPT 1260  
Db 1187 SYIISOPIIIMYSLDGKKWQTYRGNSTGTLWVFGNVDSGIGKHINFPPIIARYIRLHPT 1246  
QY 1261 HYSIRSTLWELMGCDLNSCSPLGMSKASIDAOITASSYFTNMFATWSPSKARLHLQ 1320  
Db 1247 HYSIRSTLWELMGCDLNSCSPLGMSKASIDAOITASSYFTNMFATWSPSKARLHLQ 1306  
QY 1321 RSNARPOVNNPKEMLVQDFQKTMKVTGVTQGVKSLLTSMYVKEFLISSQDGHQWTLF 1380  
Db 1307 RSNARPOVNNPKEMLVQDFQKTMKVTGVTQGVKSLLTSMYVKEFLISSQDGHQWTLF 1366  
QY 1381 FQNGKVKVQGNQDSFTPVVNSLDPLLRILYLRHQPQSWHQIALRMEVLGCEAQDLY 1438  
Db 1367 FQNGKVKVQGNQDSFTPVVNSLDPLLRILYLRHQPQSWHQIALRMEVLGCEAQDLY 1424

RESULT 14  
ABG92541  
ID ABG92541 standard; Protein; 1447 AA.  
XX AC ABG92541;  
XX DT 19-NOV-2002 (first entry)  
XX DE 5Arg B-domain-deleted-Factor VIII (FVIII).  
XX KW Human; alpha-galactosidase; Factor VIII; Factor IX;  
XX KW gene therapy.  
XX OS Homo sapiens.  
XX OS Synthetic.  
XX PN WO200264799-A2.  
XX PN 22-AUG-2002.  
XX PD 11-OCT-2001; 2001WO-US42655.  
XX PF 11-OCT-2000; 2000US-0686497.  
XX PR (TRAN-) TRANSKARYOTIC THERAPIES INC.  
XX PA Seldon RF, Miller AM, Treco DS;  
XX PI WPI: 2002-627600/67.  
XX DR N-PSDB; ABS68000.  
XX PT New synthetic nucleic acid sequence involving a continuous stretch of a  
PT least 150 common codons, useful for expressing mammalian, preferably

PT human proteins e.g. alpha-galactosidase or Factor VIII or IX or for  
PT gene therapy -  
XX  
PS Example 1; Figure 9; 115pp; English.  
XX  
CC The invention relates to a synthetic nucleic acid sequence, where at  
CC least one non-common or less-common codon is replaced with a common  
CC codon. The synthetic nucleic acid is useful for expressing mammalian,  
CC preferably human proteins e.g. alpha-galactosidase or Factor VIII or IX  
CC or for gene therapy. The synthetic nucleic acid allows precise dosing and  
CC reduces treatment costs. It is simple to apply in treating patients and  
CC is curative (one gene therapy treatment has the potential to last a  
CC patient's lifetime. ABG92540-ABG92541 represent Factor VIII and Factor  
XX IX synthetic amino acid sequences of the invention.  
SQ Sequence 1447 AA;  
Query Match 98.7%; Score 7588; DB 23; Length 1447;  
Best Local Similarity 99.0%; Pred. No. 0;  
Matches 1423; Conservative 1; Mismatches 4; Indels 10; Gaps 1;  
QY 1 ATRRYLGAVELSWDYMOSDLGELPVDARFPFRVPEKSPFNSTVYVKKTLFVEFTVHLFN 60  
Db 20 ATRRYLGAVELSWDYMOSDLGELPVDARFPFRVPEKSPFNSTVYVKKTLFVEFTVHLFN 79  
QY 61 TAKPRPPMGLGPTIOAEVYDVTVTITLKNMASHPVSLHAGVSVYKASEGAEYDDQTSQ 120  
Db 80 TAKPRPPMGLGPTIOAEVYDVTVTITLKNMASHPVSLHAGVSVYKASEGAEYDDQTSQ 139  
QY 121 REKEDKVPFGSSHTYVQVLKENGPMASDPLCLTYSYLSHVDLYKDLNSGLIGALLVCR 180  
Db 140 REKEDKVPFGSSHTYVQVLKENGPMASDPLCLTYSYLSHVDLYKDLNSGLIGALLVCR 199  
QY 181 EGS�AKEKTQTLHKILLFAVFEDEKSWHSETKNSLMQDRDAASARAPMHTVGVYNR 240  
Db 200 EGS�AKEKTQTLHKILLFAVFEDEKSWHSETKNSLMQDRDAASARAPMHTVGVYNR 259  
QY 241 SLPLGLIGCHRSVYWHVIGMGTTPVHSIFLEGHTFLVRNHRQASLEISPIFLTAQTLL 300  
Db 260 SLPLGLIGCHRSVYWHVIGMGTTPVHSIFLEGHTFLVRNHRQASLEISPIFLTAQTLL 319  
QY 301 MDLQGLFLFCHTSSHQHDMGEAYKVDSCPEBPQLRMKNNEAEYDDDLTDSMDVVRF 360  
Db 320 MDLQGLFLFCHTSSHQHDMGEAYKVDSCPEBPQLRMKNNEAEYDDDLTDSMDVVRF 379  
QY 361 DDNSPSFIQIRSVAKKPKTWVHYIAAEEDWDVAPLAPDDSYKSOYLNNGPORIG 420  
Db 380 DDNSPSFIQIRSVAKKPKTWVHYIAAEEDWDVAPLAPDDSYKSOYLNNGPORIG 439  
QY 421 RYKVKVRFMAYTDETFKTREAIQHESGILGPLLYGEVGDTLIIIFKNQASRPNIYPHGI 480  
Db 440 RYKVKVRFMAYTDETFKTREAIQHESGILGPLLYGEVGDTLIIIFKNQASRPNIYPHGI 499  
QY 481 TDVRLYLRRLPKGVKHLKDFPILGEPFKYKWTVTVEDGPTKSDPRCLTRYSSPVNME 540  
Db 500 TDVRLYLRRLPKGVKHLKDFPILGEPFKYKWTVTVEDGPTKSDPRCLTRYSSPVNME 559  
QY 541 RDLASGLIGPLLI CYKESVDQRGNQIMSDKRNVLFSVFDENRSWLTENIQFLPNPAG 600  
Db 560 RDLASGLIGPLLI CYKESVDQRGNQIMSDKRNVLFSVFDENRSWLTENIQFLPNPAG 619  
QY 601 VOLEDPEFQASNIHMSINGVYFDSQLSVCLHEVAYWYLSIGAQTDPLSFVSGVTFKH 660  
Db 620 VOLEDPEFQASNIHMSINGVYFDSQLSVCLHEVAYWYLSIGAQTDPLSFVSGVTFKH 679  
QY 661 KMWYEDTLTLPFPGSETVPMSENFGMLILGCHNSDFRNRGMTALLKVSSCDKNTGDYFE 720  
Db 680 KMWYEDTLTLPFPGSETVPMSENFGMLILGCHNSDFRNRGMTALLKVSSCDKNTGDYFE 739  
QY 721 DSYVEDISAYLLSKNNAIEPRSPQNPVLKXHQREITRTLQSDQEEIDYDDTISVENKK 780  
Db 740 DSYVEDISAYLLSKNNAIEPRSPQNPVLKXHQREITRTLQSDQEEIDYDDTISVENKK 789



Db 662 KMVYEDTJTLRPFSGETVFMSENFGMLWILGCHNSDFRNRGMTALLKVSSCDKNTGDIYE 721  
QY 721 DSYEDISAYLLSKNNAIEPRSPQNPVVLKRHRQREITRTTQSDQOEEDYDDTISVEMKK 780  
Db 722 DSYEDISAYLLSKNNAIEP-----REITRTTQSDQOEEDYDDTISVEMKK 767  
QY 781 EDFDIYDEENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNRRAQSGSVPOFKKV 840  
Db 768 EDFDIYDEENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNRRAQSGSVPOFKKV 827  
QY 841 FOEFTDGSFTQPLYGELNEHGLAGPYIRAEVEDNIMVTFRNOASRPVSYSSLSIYEE 900  
Db 828 FOEFTDGSFTQPLYGELNEHGLAGPYIRAEVEDNIMVTFRNOASRPVSYSSLSIYEE 887  
QY 901 DORQGAEPFRKNEFKNTKTYFMKVQHMAPTKDFCKAWAYFSDVDLEKDVHSGLIGP 960  
Db 888 DORQGAEPFRKNEFKNTKTYFMKVQHMAPTKDFCKAWAYFSDVDLEKDVHSGLIGP 947  
QY 961 LLVCHTNTLNPAGHQVTVQEFALFTTFDETQSWYFTENMERNCRAPCNIQMEDPTFKE 1020  
Db 948 LLVCHTNTLNPAGHQVTVQEFALFTTFDETQSWYFTENMERNCRAPCNIQMEDPTFKE 1007  
QY 1021 NYRFRHAINGYIMDTLPGLVMAODQIRWYLLSGNSNENIHSIFSGHVFTVRKKEBKMA 1080  
Db 1008 NYRFRHAINGYIMDTLPGLVMAODQIRWYLLSGNSNENIHSIFSGHVFTVRKKEBKMA 1067  
QY 1081 LYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPPLGMAHGIRD 1140  
Db 1068 LYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPPLGMAHGIRD 1127  
QY 1141 FOITASQGYQWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMIIHGIKTOGAROKFS 1200  
Db 1128 FOITASQGYQWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMIIHGIKTOGAROKFS 1187  
QY 1201 SLYISQFIIMYSLDGKKWOTYRGNSTGLMVFFGNVDSSGIKHNFNPPIIARYIRLHPT 1260  
Db 1188 SLYISQFIIMYSLDGKKWOTYRGNSTGLMVFFGNVDSSGIKHNFNPPIIARYIRLHPT 1247  
QY 1261 HYSIRSTLRMELMGCDLNSCMLPGMESKAISDAQITASSYFTNMFPATWSPSKARHLQ 1320  
Db 1248 HYSIRSTLRMELMGCDLNSCMLPGMESKAISDAQITASSYFTNMFPATWSPSKARHLQ 1307  
QY 1321 RSNARWPQVNNPKEMLQVDFQKTMKVTGTTQGVKSLLTSMYVKEFLISSSQDGHQWTLF 1380  
Db 1308 RSNARWPQVNNPKEMLQVDFQKTMKVTGTTQGVKSLLTSMYVKEFLISSSQDGHQWTLF 1367  
QY 1381 FONGKVKVFGNQDSFTPVVNSLDPLLRILRIHPQSWHQAIALRMEVLGCEAODLY 1438  
Db 1368 FONGKVKVFGNQDSFTPVVNSLDPLLRILRIHPQSWHQAIALRMEVLGCEAODLY 1425

Search completed: December 9, 2003, 16:53:08  
Job time : 67 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 9, 2003, 16:52:06 ; Search time 22 Seconds  
(without alignments)  
2765.592 Million cell updates/sec

Title: US-10-006-091-1  
Perfect score: 7691  
Sequence: 1 ATRRYLGAVELSWDMQSD.....VWQHIALRMEVLGCEAQLY 1438

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pap.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pap.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pap.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pap.\*  
5: /cgn2\_6/ptodata/1/iaa/PTUS\_COMB.pap.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7691	100.0	1438	4	US-09-209-916-1
2	7674	99.8	1471	1	US-08-683-839B-3
3	7434	96.7	1661	2	US-08-882-083-2
4	7434	96.7	1661	2	US-08-558-107-2
5	7434	96.7	1661	3	US-09-243-539-2
6	7234	94.1	2332	1	US-07-864-004B-4
7	7234	94.1	2332	1	US-08-251-937A-4
8	7234	94.1	2332	1	US-08-212-133A-2
9	7234	94.1	2332	1	US-08-474-503-2
10	7234	94.1	2332	1	US-08-670-707A-2
11	7234	94.1	2332	3	US-09-037-601-2
12	7234	94.1	2332	4	US-09-315-179-2
13	7234	94.1	2332	4	US-09-523-656-2
14	7234	94.1	2332	5	PCT-US93-03275-4
15	7234	94.1	2332	5	PCT-US94-13200-6
16	7234	94.1	2351	1	US-08-121-202-2
17	7227	94.0	2351	1	US-08-366-851A-2
18	7225	93.9	2351	6	542260-1
19	7224	93.9	2332	1	US-08-276-594A-2
20	7221	93.9	2351	6	5171844-2
21	7195	93.6	2332	3	US-09-324-867-3
22	6560	85.3	1467	4	US-09-523-656-38
23	6489	84.4	1443	2	US-08-670-707A-39
24	6489	84.4	1443	3	US-09-037-601-39
25	6489	84.4	1443	4	US-09-315-179-39
26	6272	81.5	2343	3	US-09-324-867-2
27	6227	81.0	2133	2	US-08-670-707A-37

28	6227	81.0	2133	3	US-09-037-601-37	Sequence 37, Appl
29	6227	81.0	2133	4	US-09-315-179-37	Sequence 37, Appl
30	6227	81.0	2133	4	US-09-523-656-30	Sequence 30, Appl
31	6192	80.5	2319	1	US-08-212-133A-8	Sequence 8, Appl
32	6192	80.5	2319	1	US-08-474-503-6	Sequence 6, Appl
33	6192	80.5	2319	2	US-08-670-707A-6	Sequence 6, Appl
34	6192	80.5	2319	3	US-09-037-601-6	Sequence 6, Appl
35	6192	80.5	2319	4	US-09-315-179-6	Sequence 6, Appl
36	6192	80.5	2319	4	US-09-523-656-28	Sequence 28, Appl
37	6192	80.5	2319	5	PCT-US94-13200-6	Sequence 6, Appl
38	6187.5	80.5	2115	3	US-09-324-867-5	Sequence 5, Appl
39	6147	79.9	2304	3	US-09-121-202-4	Sequence 4, Appl
40	2444.5	31.8	541	1	US-08-121-202-4	Sequence 4, Appl
41	2369	30.8	2183	3	US-08-746-111-5	Sequence 5, Appl
42	1726	22.4	368	1	US-08-212-133A-6	Sequence 6, Appl
43	1726	22.4	368	1	US-08-474-503-4	Sequence 4, Appl
44	1726	22.4	368	2	US-08-670-707A-4	Sequence 4, Appl
45	1726	22.4	368	3	US-09-037-601-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1  
US-09-209-916-1  
; Sequence 1, Application US/09209916  
; Patent No. 6358703  
; GENERAL INFORMATION:  
; APPLICANT: Cho, Myung-Sam  
; APPLICANT: Chan, Sham-Yuen  
; APPLICANT: Kelsey, William  
; APPLICANT: Yee, Helena  
; TITLE OF INVENTION: Expression System for Factor VIII  
; FILE REFERENCE: MSB-7255  
; CURRENT APPLICATION NUMBER: US/09/209,916  
; CURRENT FILING DATE: 1998-12-10  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 1438  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Derived from  
; OTHER INFORMATION: human factor VIII sequence  
US-09-209-916-1

Query Match	100.0%	Score	7691	DB	4	Length	1438
Best Local Similarity	100.0%	Pred. No.	0	Mismatches	0	Indels	0
Matches	1438	Conservative	0	Gaps	0		
Qy	1	ATRRYILGAVELSWDMQSDLGELPVDARPPRPVKSPFNNTSVVYKKTFLVFEFTVHLFN	60				
Db	1	ATRRYILGAVELSWDMQSDLGELPVDARPPRPVKSPFNNTSVVYKKTFLVFEFTVHLFN	60				
Qy	61	IAKPRPPMGLGPTIQAEVYDVTVTITLKNMASHPVSLHVGVSYWKASGAEYDDQTSQ	120				
Db	61	IAKPRPPMGLGPTIQAEVYDVTVTITLKNMASHPVSLHVGVSYWKASGAEYDDQTSQ	120				
Qy	121	REKEDDKVFPGGSHYVQVVKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR	180				
Db	121	REKEDDKVFPGGSHYVQVVKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR	180				
Qy	181	EGSLAKEKTOFLHKFILLFAVFDGKSWHSKTNLSMDQDRDAASARAWPKMHTVNGYVR	240				
Db	181	EGSLAKEKTOFLHKFILLFAVFDGKSWHSKTNLSMDQDRDAASARAWPKMHTVNGYVR	240				
Qy	241	SLPGLIGCHRSVYVHVGMTTPEVHSIFLEGHTFLVRNHRQASLEISPTITFLTAQTL	300				
Db	241	SLPGLIGCHRSVYVHVGMTTPEVHSIFLEGHTFLVRNHRQASLEISPTITFLTAQTL	300				
Qy	301	MDLGQFLFCHSHSHQHDGMAYVKVDSCPREPQPMKNNEAEYDDDLTDSMDVVRP	360				
Db	301	MDLGQFLFCHSHSHQHDGMAYVKVDSCPREPQPMKNNEAEYDDDLTDSMDVVRP	360				

Db 301 MDLQGLLFCHISSHQHDGMEAYVKVDSCEPFPQLRMKNNEEAEDYDDDLTDSMDVVRP 360  
QY 361 DDNSPSFTQIRSVAKKPKTWVHYTAAREEDWDAPLVLAAPDRSRYKSOYLNNGPQIRG 420  
Db 361 DDNSPSFTQIRSVAKKPKTWVHYTAAREEDWDAPLVLAAPDRSRYKSOYLNNGPQIRG 420  
QY 421 RYKVKRPMAYTDFTKTREAIOHESGILGLPLYGEVGTLLIIFKNQASRPYNTYPHGI 480  
Db 421 RYKVKRPMAYTDFTKTREAIOHESGILGLPLYGEVGTLLIIFKNQASRPYNTYPHGI 480  
QY 481 TVRPLYSRLPKGVKHLKDFILPGEIFKYKWTVTVEDGPKSDPRCLTRYSSFFVNM 540  
Db 481 TVRPLYSRLPKGVKHLKDFILPGEIFKYKWTVTVEDGPKSDPRCLTRYSSFFVNM 540  
QY 541 RLASGLIGPLLI CYKESVDQGNQIMSDKRNVLIFSVDENRWSWLTENIQRFNPAG 600  
Db 541 RLASGLIGPLLI CYKESVDQGNQIMSDKRNVLIFSVDENRWSWLTENIQRFNPAG 600  
QY 601 VQLEDEPFOASNMHSINGVYFDSLSQVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKH 660  
Db 601 VQLEDEPFOASNMHSINGVYFDSLSQVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKH 660  
QY 661 KMVEDTLTLFPFSGETVPMENPGLWILGCHNSDFRNGMTALLKVSSCDKNTGDYIE 720  
Db 661 KMVEDTLTLFPFSGETVPMENPGLWILGCHNSDFRNGMTALLKVSSCDKNTGDYIE 720  
QY 721 DSYEDISAVLLSKNNAIEPRSPNPVLKRQREITRTTLOSQDREIDYDITISVEMKK 780  
Db 721 DSYEDISAVLLSKNNAIEPRSPNPVLKRQREITRTTLOSQDREIDYDITISVEMKK 780  
QY 781 EDFDIYDENOSPRSFOKTRHYFIAAVERLWDYGMSSSPHYLRNRAQSGVPQPKV 840  
Db 781 EDFDIYDENOSPRSFOKTRHYFIAAVERLWDYGMSSSPHYLRNRAQSGVPQPKV 840  
QY 841 FOEFTDGSFTQPLRGELNEHGLLGPYIRAEVEDNIMVTFRNOASRPYSFSSLSIYEE 900  
Db 841 FOEFTDGSFTQPLRGELNEHGLLGPYIRAEVEDNIMVTFRNOASRPYSFSSLSIYEE 900  
QY 901 DORQGAEPKRFVKNETKTYFMKVOHMAPTKDEDFCKAWYFSDVLEKDVHSLGIP 960  
Db 901 DORQGAEPKRFVKNETKTYFMKVOHMAPTKDEDFCKAWYFSDVLEKDVHSLGIP 960  
QY 961 LLVCHTNTLPAHGRQVTVQEFALFTTIFDETSWYFTENMERNCRAPCNIOEDPTFK 1020  
Db 961 LLVCHTNTLPAHGRQVTVQEFALFTTIFDETSWYFTENMERNCRAPCNIOEDPTFK 1020  
QY 1021 NTRFHAINGYIMDTLGLVMAQDQRIWYLLSGNSNENIHSIFSGHVTVRKKEKYKA 1080  
Db 1021 NTRFHAINGYIMDTLGLVMAQDQRIWYLLSGNSNENIHSIFSGHVTVRKKEKYKA 1080  
QY 1081 LYNLYPGVFTVEMLPKAGIWRVECLIGELHAGMSTLFLVYSNKCOTPLGNASGHIRD 1140  
Db 1081 LYNLYPGVFTVEMLPKAGIWRVECLIGELHAGMSTLFLVYSNKCOTPLGNASGHIRD 1140  
QY 1141 FOITASQYQWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMIHIGIKTOGARQFS 1200  
Db 1141 FOITASQYQWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMIHIGIKTOGARQFS 1200  
QY 1201 SLVISOPIIMSYLDGKKWQYTRNGSTGLTMVFFGNVDSSGINKINFPPIIARYIRLHPT 1260  
Db 1201 SLVISOPIIMSYLDGKKWQYTRNGSTGLTMVFFGNVDSSGINKINFPPIIARYIRLHPT 1260  
QY 1261 HYSIRSTRMELMCDLNSCSMPLGMSKAI SDAQITASSYFTNMATWSPSKARLHLQ 1320  
Db 1261 HYSIRSTRMELMCDLNSCSMPLGMSKAI SDAQITASSYFTNMATWSPSKARLHLQ 1320  
QY 1321 RSNARVPQVNPKEWLQVDFQKTMKVTVGTITQGVKSLTSMYKFEFLISSQDGHQWTLF 1380  
Db 1321 RSNARVPQVNPKEWLQVDFQKTMKVTVGTITQGVKSLTSMYKFEFLISSQDGHQWTLF 1380  
QY 1381 FQNGKVKVFGNQDQSFPTVNVNSLDPLLTLYRLRHPQSVHQIALRMEVLGCAQDILY 1438  
Db 1381 FQNGKVKVFGNQDQSFPTVNVNSLDPLLTLYRLRHPQSVHQIALRMEVLGCAQDILY 1438

## RESULT 2

US-08-839B-3  
; Sequence 3, Application US/08683839B  
; Patent No. 574326

## ; GENERAL INFORMATION:

; APPLICANT: Ili, Charles . R. et al.  
; TITLE OF INVENTION: Use of Viral Cis-Acting Post-Transcriptional  
; TITLE OF INVENTION: Regulatory Sequences to Increase Expression of  
; TITLE OF INVENTION: Introns Genes Containing Near-Consensus Splice Sites  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: LAHIVE & COCKFIELD  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA

; ZIP: 02109-1875

## ; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/683.839B  
; FILING DATE: 11-MARCH-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:

## ; FILING DATE:

; ATTORNEY/AGENT INFORMATION:  
; NAME: Remillard, Jane E.  
; REGISTRATION NUMBER: 38,872  
; REFERENCE/DOCKET NUMBER: TTI-138  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)227-7400  
; TELEFAX: (617)227-5941

## ; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1471 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein

US-08-683-839B-3

Query Match 99.8%; Score 7674; DB 1; Length 1471;

Best Local Similarity 99.0%; Pred. No. 0;

Matches 1438; Conservative 0; Mismatches 0; Indels 14; Gaps 1;

QY 1 ATRRYLGAVELSWDYMQSDLGELPVDARPPRPVKSPFPNTSVVYKKTFLVEFTVHLFN 60  
Db 20 ATRRYLGAVELSWDYMQSDLGELPVDARPPRPVKSPFPNTSVVYKKTFLVEFTVHLFN 79  
QY 61 IAKPRPPMGLLGPTIOAEVYTVITLKNMASHPVSLHAGVSVTWKASEGAEYDDQTSQ 120  
Db 80 IAKPRPPMGLLGPTIOAEVYTVITLKNMASHPVSLHAGVSVTWKASEGAEYDDQTSQ 139  
QY 121 RKEDDKVPFGSGHTVVMQVLKENGPMASDPLCLTYSLYSHVDLVKLNLSGLIGALLVCR 180  
Db 140 RKEDDKVPFGSGHTVVMQVLKENGPMASDPLCLTYSLYSHVDLVKLNLSGLIGALLVCR 199  
QY 181 EGS LAKEKTQTLHKFTLLFAVDFEGKSWHSEKNSLMQDRDAASARAWPKMFTVNGYVNR 240  
Db 200 EGS LAKEKTQTLHKFTLLFAVDFEGKSWHSEKNSLMQDRDAASARAWPKMFTVNGYVNR 259  
QY 241 SLPGILGCHRKSVYMHVIGMGTTPVHSIFLEGHTFLVRNHRQASLEISPIFLTAQTLL 300  
Db 260 SLPGILGCHRKSVYMHVIGMGTTPVHSIFLEGHTFLVRNHRQASLEISPIFLTAQTLL 319  
QY 301 MDLQGLLFCHISSHQHDGMEAYVKVDSCEPFPQLRMKNNEEAEDYDDDLTDSMDVVRP 360  
Db 320 MDLQGLLFCHISSHQHDGMEAYVKVDSCEPFPQLRMKNNEEAEDYDDDLTDSMDVVRP 379

QY 361 DDNSPSFIQIRSVAKKPKTWYHIAAEEEDMDYAPLVLPADDRSKSYQLNNGPORIG 420  
DB 380 DDNSPSFIQIRSVAKKPKTWYHIAAEEEDMDYAPLVLPADDRSKSYQLNNGPORIG 439  
QY 421 RYKVKVFMAYTDETFKTRAIQESGILGPLLGEVGDITLLIFKNQASRPYNIYPHGI 480  
DB 440 RYKVKVFMAYTDETFKTRAIQESGILGPLLGEVGDITLLIFKNQASRPYNIYPHGI 499  
QY 481 TDVRPLYSRRLPKGVKHLKDPFLLPGHIFPKYKWTVTVEDGFTKSDPCLTRYSSFYNME 540  
DB 500 TDVRPLYSRRLPKGVKHLKDPFLLPGHIFPKYKWTVTVEDGFTKSDPCLTRYSSFYNME 559  
QY 541 RDLASGLIGPLLICYKESVDORGNOIMSDKRNVLFSVFDENRSWYLTENIQRLPNPAG 600  
DB 560 RDLASGLIGPLLICYKESVDORGNOIMSDKRNVLFSVFDENRSWYLTENIQRLPNPAG 619  
QY 601 VOLEDEPFOASINMHSINGVYVDSIQLSVCLHEVAYWYILSIGAQTDPLSVFFSGYTFKH 660  
DB 620 VOLEDEPFOASINMHSINGVYVDSIQLSVCLHEVAYWYILSIGAQTDPLSVFFSGYTFKH 679  
QY 661 KMVYEDTILTFPFSGETVFMENPGLMILGCHNSDFNRNGMTALLKVSSCDKNTGDIYE 720  
DB 680 KMVYEDTILTFPFSGETVFMENPGLMILGCHNSDFNRNGMTALLKVSSCDKNTGDIYE 739  
QY 721 DSYEDISAYLLSKNNAIEPRFSQN-----PPVLKXHQREITRTTLOSQOE 766  
DB 740 DSYEDISAYLLSKNNAIEPRFSQNSRHPSTROKQFNATPPVLKXHQREITRTTLOSQOE 799  
QY 767 EIDYDDTISVEMKKEDEDIYDEENOSPRSKTRHYFIAAVERLMDYGMSSPHVLNR 826  
DB 800 EIDYDDTISVEMKKEDEDIYDEENOSPRSKTRHYFIAAVERLMDYGMSSPHVLNR 859  
QY 827 RAQSGSVPOFKVVFQFETDGSFTQPLYRGELNEHLGLLGYIRAEVEDNIMVTFRNOAS 886  
DB 860 RAQSGSVPOFKVVFQFETDGSFTQPLYRGELNEHLGLLGYIRAEVEDNIMVTFRNOAS 919  
QY 887 RPSYFYSLSIYZEDQOGAEPKRNKVPKNETKTYFMKVQHMAPTKDEPCKAWAYPSD 946  
DB 920 RPSYFYSLSIYZEDQOGAEPKRNKVPKNETKTYFMKVQHMAPTKDEPCKAWAYPSD 979  
QY 947 VLEKDVHSGILGPLLVCHTNTLPAHGRQVTVQEPALFFTIDETKSWYFTEWNERCR 1006  
DB 980 VLEKDVHSGILGPLLVCHTNTLPAHGRQVTVQEPALFFTIDETKSWYFTEWNERCR 1039  
QY 1007 APCNIQMEDPTFKENYRPHALNGYIMDTPLGLVMAQDQIRWYLLSMGNSNENIHSIFSG 1066  
DB 1040 APCNIQMEDPTFKENYRPHALNGYIMDTPLGLVMAQDQIRWYLLSMGNSNENIHSIFSG 1099  
QY 1067 HVFTVRKKEEYKMALXNLYPGVFTVEMLPKAGIWRVECLIGELHAGMSTLFLVYSNK 1126  
DB 1100 HVFTVRKKEEYKMALXNLYPGVFTVEMLPKAGIWRVECLIGELHAGMSTLFLVYSNK 1159  
QY 1127 COTPLGMASGHIRFOITASGOYGQWAPKLARLHYSGSINAWSTKEPFSWKVLDLAPMI 1186  
DB 1160 COTPLGMASGHIRFOITASGOYGQWAPKLARLHYSGSINAWSTKEPFSWKVLDLAPMI 1219  
QY 1187 IHGIKTOGARQKFSLSYISQFIIMYSLDGKKWQYRGNSTGTLMVFFGNVDSGGIKNIF 1246  
DB 1220 IHGIKTOGARQKFSLSYISQFIIMYSLDGKKWQYRGNSTGTLMVFFGNVDSGGIKNIF 1279  
QY 1247 NPPIIARIYRLHPTHTYSIRSTRMLMGCNDLNSCMPLGMSKASDAQITASSYFTNMF 1306  
DB 1280 NPPIIARIYRLHPTHTYSIRSTRMLMGCNDLNSCMPLGMSKASDAQITASSYFTNMF 1339  
QY 1307 ATWSPSKARLHLQGRSNARPOVNNPKWLQVDFQKTMKVTCTVTCQVKSLLTSMYKKEF 1366  
DB 1340 ATWSPSKARLHLQGRSNARPOVNNPKWLQVDFQKTMKVTCTVTCQVKSLLTSMYKKEF 1399  
QY 1367 LISSSDQGHQWTLTFQNGKVKYFQGNQDSFTPVNSLDPPLLTRYLRIHPQSWHQAIALR 1426  
DB 1400 LISSSDQGHQWTLTFQNGKVKYFQGNQDSFTPVNSLDPPLLTRYLRIHPQSWHQAIALR 1459

QY 1427 MEVLGCEAQDLY 1438  
DB 1460 MEVLGCEAQDLY 1471  
RESULT 3  
US-08-882-083-2  
; Sequence 2, Application US/08882083  
; Patent No. 5863292  
; GENERAL INFORMATION:  
; APPLICANT: VORBERG, Johannes J.  
; TITLE OF INVENTION: HYBRID PROTEINS WITH MODIFIED ACTIVITY  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA: US/08/882,083  
; FILING DATE:  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/558,107  
; FILING DATE: 13-NOV-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: ISACSON, John P.  
; REGISTRATION NUMBER: 33,715  
; REFERENCE/DOCKET NUMBER: 30472/212  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1661 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-882-083-2

Query Match 96.7%; Score 7434; DB 2; Length 1661;  
Best Local Similarity 86.0%; Pred. No. 0;  
Matches 1418; Conservative 4; Mismatches 10; Indels 216; Gaps 4;

QY 1 ATRRYLGAVELSWDYMQSDIGELPVDARFPVRPKSPFNTSVVYKKTFLVFEFVHLFN 60  
DB 20 ATRRYLGAVELSWDYMQSDIGELPVDARFPVRPKSPFNTSVVYKKTFLVFEFVHLFN 79  
QY 61 IAKPRPPMGLLGPTIOAEVYDVTIVITLKNASHPVSLHVGVSYWKASGAEYDDQTSQ 120  
DB 80 IAKPRPPMGLLGPTIOAEVYDVTIVITLKNASHPVSLHVGVSYWKASGAEYDDQTSQ 139  
QY 121 REKEDDKVFPKGSSTYVWQVLEKNGPMSDPLCLTYSYVLSHVDLVKDLNSGLIGALLVCR 180  
DB 140 REKEDDKVFPKGSSTYVWQVLEKNGPMSDPLCLTYSYVLSHVDLVKDLNSGLIGALLVCR 199  
QY 181 EGS�AKEKTOTLHKPILLFVAFVDEKGSWHSSTKNSLMQDRDAASARAWPKMHTVNGYVNR 240  
DB 200 EGS�AKEKTOTLHKPILLFVAFVDEKGSWHSSTKNSLMQDRDAASARAWPKMHTVNGYVNR 259  
QY 241 SLPLGIGCHRSKVYWHVIGMGTTPPEVHSIFLEGTFLVRNHRQASLEISPTTFTTAQTLL 300  
DB 260 SLPLGIGCHRSKVYWHVIGMGTTPPEVHSIFLEGTFLVRNHRQASLEISPTTFTTAQTLL 319  
QY 301 MDLQOFLFLFCHISSHQHDGMEAYVKVDSCPREPQLRMKNNEAEDYDDDLTDSMDVVRP 360

Db 320 MDLGGFLFCHISSHQHDGMEAYVVDSCPEBPQRLMKNEEAEDYDDDLTSEMDVYVF 379  
Qy 361 DDNSPSTQTSVAKKPKTWVHYIAAEEEDWDYAPLVLAPDDRSYKQYLNNGPQRIQ 420  
Db 380 DDNSPSTQTSVAKKPKTWVHYIAAEEEDWDYAPLVLAPDDRSYKQYLNNGPQRIQ 439  
Qy 421 RYKVRFWAYTDETFKREATQHSIGLIPLLYGEVGDLLIIFKNQASRPYNIYPHGI 480  
Db 440 RYKVRFWAYTDETFKREATQHSIGLIPLLYGEVGDLLIIFKNQASRPYNIYPHGI 499  
Qy 481 TDVRLYRRLPKGVKHLKDFILPCEIPKYKWTVTVEDGPKSDPRCLTRYSSFYVME 540  
Db 500 TDVRLYRRLPKGVKHLKDFILPCEIPKYKWTVTVEDGPKSDPRCLTRYSSFYVME 559  
Qy 541 RLASGLIPLLI CYKESVDQGNQIMSDKRNVLFSVFDENRSMWLTENIORFLPNPAG 600  
Db 560 RLASGLIPLLI CYKESVDQGNQIMSDKRNVLFSVFDENRSMWLTENIORFLPNPAG 619  
Qy 601 VOLEDPEFOASNIMHSINGVYFDSLSQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYTPKH 660  
Db 620 VOLEDPEFOASNIMHSINGVYFDSLSQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYTPKH 679  
Qy 661 KMYEDTTLFPFSGETVMSMENPGLWILGCHNSDFRNRGMTALLKVSSC---DKYTCGD 717  
Db 680 KMYEDTTLFPFSGETVMSMENPGLWILGCHNSDFRNRGMTALLKVSSCIPBGEEDDD 739  
Qy 718 YX-----EYSYEDISAYLLSKNAIEPRSF----- 742  
Db 740 YLDLEKIFSEDDYDI-----VDSLEPRFSQNSRHPSTROKQFNATTIPENDIEKT 793  
Qy 743 ----- 742  
Db 794 DPWFARTPMPKIQNVSSDILLMLLQSPTPHGLSLSLQEAKEYTFSDPSFGAIDSN 853  
Qy 743 ----- 742  
Db 854 SLSEMTFRPOLHSGDMVTPESGLQLRLNBKLGTTADPLAWNHYGTQIPEEKWSOE 913  
Qy 743 -----SONPPVLK 750  
Db 914 KSPEKTAFKKOTILSNACESHATAANEQONKEPFIETWAKQORTERLCSQNPVLK 973  
Qy 751 RQREITRTLOSQDEIDYDDTISVEMKKEDDIYDEENOSPRSFOKKTTHYFIAAYE 810  
Db 974 RQREITRTLOSQDEIDYDDTISVEMKKEDDIYDEENOSPRSFOKKTTHYFIAAYE 1033  
Qy 811 RLWDYQWSSPHVLRNRAQSGSVPPQKVVQFQFTDGSFTQPLYRGELNEHLGLLPYIR 870  
Db 1034 RLWDYQWSSPHVLRNRAQSGSVPPQKVVQFQFTDGSFTQPLYRGELNEHLGLLPYIR 1093  
Qy 871 AEVEDNIMVTRNQAQSRPSFYSSLSIYEDDQGAEPKRNFKVPNETKTYFWKVQHMA 930  
Db 1094 AEVEDNIMVTRNQAQSRPSFYSSLSIYEDDQGAEPKRNFKVPNETKTYFWKVQHMA 1153  
Qy 931 PTKDEFCKAWAYFSDVLEKDVHSGLIGPLLYCHTNTLNPAHGRQVTVQEPALFTTFD 990  
Db 1154 PTKDEFCKAWAYFSDVLEKDVHSGLIGPLLYCHTNTLNPAHGRQVTVQEPALFTTFD 1213  
Qy 991 ETKSWYFTENMERNCRAPCMQIOMEDPTFKENYRFAHNGYIMDTLGLVMAQORIRWYL 1050  
Db 1214 ETKSWYFTENMERNCRAPCMQIOMEDPTFKENYRFAHNGYIMDTLGLVMAQORIRWYL 1273  
Qy 1051 LSWGNSNETHSHFSGHVFTRKKEEYKMAVLYNPGVFTVEMLPKAGIWRVECLIGE 1110  
Db 1274 LSWGNSNETHSHFSGHVFTRKKEEYKMAVLYNPGVFTVEMLPKAGIWRVECLIGE 1333  
Qy 1111 HLHAGNSTLFLVYSNCKQTPGLMASGHIRDFQITASQYQWAPKARLHYSGSINAWST 1170  
Db 1334 HLHAGNSTLFLVYSNCKQTPGLMASGHIRDFQITASQYQWAPKARLHYSGSINAWST 1393  
Qy 1171 KEFPSWIKVDLLAPMIHGIKTQAGOKFSSLYISQFIIMYSLDGKKWQYRGNSTGTL 1230

Db 1394 KEFPSWIKVDLLAPMIHGIKTQAGOKFSSLYISQFIIMYSLDGKKWQYRGNSTGTL 1453  
Qy 1231 VFGNVDSSGKHNIFNPPIIARIYIRLHPHYSTRIRLMELMGCDLNSCNPGLMESKA 1290  
Db 1454 VFGNVDSSGKHNIFNPPIIARIYIRLHPHYSTRIRLMELMGCDLNSCNPGLMESKA 1513  
Qy 1291 ISDAQITASSYFTWMEATWSPSKARLHIQGRSNARWPOVNNPKWLOVDFOKTKVTCVT 1350  
Db 1514 ISDAQITASSYFTWMEATWSPSKARLHIQGRSNARWPOVNNPKWLOVDFOKTKVTCVT 1573  
Qy 1351 TQGVKSLLTSMYVKEFLISSQDGHQWTLFFQNGKVKVFOGNQDSFPPVNSLDPPLLTR 1410  
Db 1574 TQGVKSLLTSMYVKEFLISSQDGHQWTLFFQNGKVKVFOGNQDSFPPVNSLDPPLLTR 1633  
Qy 1411 YLRHPQSWHQAIALRMEVLGCEAQDLY 1438  
Db 1634 YLRHPQSWHQAIALRMEVLGCEAQDLY 1661

RESULT 4

US-08-558-107-2  
; Sequence 2, Application US/08558107  
; Patent No. 5910481  
; GENERAL INFORMATION:  
; APPLICANT: VOORBERG, Johannes J.  
; TITLE OF INVENTION: HYBRID PROTEINS WITH MODIFIED ACTIVITY  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/558,107  
; FILING DATE: 13-NOV-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: ISACSON, John P.  
; REGISTRATION NUMBER: 33,715  
; REFERENCE/DOCKET NUMBER: 30472/212  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 672-5300  
; TELEFAX: (202) 672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1661 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-558-107-2

Query Match 96.7%; Score 7434; DB 2; Length 1661;  
Best Local Similarity 86.0%; Pred. No. 0;  
Matches 1418; Conservative 4; Mismatches 10; Indels 216; Gaps 4;  
Qy 1 ATRRYLGAVELSWDMQSDLGELPVDARFPFRVPKSPFPNTSVVYKKTLPVEFTVHLFN 60  
Db 20 ATRRYLGAVELSWDMQSDLGELPVDARFPFRVPKSPFPNTSVVYKKTLPVEFTVHLFN 79  
Qy 61 IAKPRPPWMLLGPTTQAQVYDTVTITLKNASHPVSLHAGVSVTKASEGAYDDQTSQ 120  
Db 80 IAKPRPPWMLLGPTTQAQVYDTVTITLKNASHPVSLHAGVSVTKASEGAYDDQTSQ 139  
Qy 121 REKEDDKVPKSGSHYVVMQVLKENGPMASDPLCTYTSYLSHVDLVKDLNSGLIGALLVCR 180

Db 140 REKEDDKVPPGGSHYVWQVLKENGPMASDPLCLTYSYLSHVLDLVDLNSGLIGALLVCR 199  
QY 181 EGS LAKEKTQTLHKFTILLFAVFDGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNR 240  
Db 200 EGS LAKEKTQTLHKFTILLFAVFDGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNR 259  
QY 241 SLPLGIGCHRSYVWIVGIMGTTPVHSIFLEGHTFLVRNHRQASLEISPIITLTAQTLL 300  
Db 260 SLPLGIGCHRSYVWIVGIMGTTPVHSIFLEGHTFLVRNHRQASLEISPIITLTAQTLL 319  
QY 301 MDLQGFLLFCHISHHQHDMGMEAVKVDSCPEEPQLMKNEEAEDYDDDLTDSEMDVRF 360  
Db 320 MDLQGFLLFCHISHHQHDMGMEAVKVDSCPEEPQLMKNEEAEDYDDDLTDSEMDVRF 379  
QY 361 DDNSPSFIQIRSAVAKKPKTWVHYIAAEEEDWDYAPLVLPDDRYSKSYQLNNGPQIRIG 420  
Db 380 DDNSPSFIQIRSAVAKKPKTWVHYIAAEEEDWDYAPLVLPDDRYSKSYQLNNGPQIRIG 439  
QY 421 RYKVKYRFMYATDETFTKTRAIQHESGILGPLLYGEVGTLLIIFKNQASRPNIYPHGI 480  
Db 440 RYKVKYRFMYATDETFTKTRAIQHESGILGPLLYGEVGTLLIIFKNQASRPNIYPHGI 499  
QY 481 TDVRPLYSRRLPKGVKHLKDPFPIIPGEIFPKYKWTVTVEDGPTKSDPRCLTRYYSFFVWME 540  
Db 500 TDVRPLYSRRLPKGVKHLKDPFPIIPGEIFPKYKWTVTVEDGPTKSDPRCLTRYYSFFVWME 559  
QY 541 RDLASGLIGPLLCYKESVDQGNQIMSDKRNVLFSVFDENRSWYLTENIQRFLENPAG 600  
Db 560 RDLASGLIGPLLCYKESVDQGNQIMSDKRNVLFSVFDENRSWYLTENIQRFLENPAG 619  
QY 601 VQLEDEPFOASNTMHSINGVYVPSLQSLVCLHEVAYWYILSIGAQTDFLSVPSGTYFKH 660  
Db 620 VQLEDEPFOASNTMHSINGVYVPSLQSLVCLHEVAYWYILSIGAQTDFLSVPSGTYFKH 679  
QY 661 KMVYEDTLTLFPFSGEIVFMSMNPGLWILGCHNSDFRNGMTALLKVSSC---DKNTGD 717  
Db 680 KMVYEDTLTLFPFSGEIVFMSMNPGLWILGCHNSDFRNGMTALLKVSSCPEGEHDD 739  
QY 718 YY-----EDSYEDISAYLLSKNNAIEPRSF----- 742  
Db 740 YLDEKIFSEDDYIDI-----VDSLIEPRFSQNSRHPSTKQKPNATTIPENDIEKT 793  
QY 743 ----- 742  
Db 794 DPWFARHTPMPKIQNVSSDLMLLRQSPTPHGLSLDLQEAKEYETPSDDPSCAIDSN 853  
QY 743 ----- 742  
Db 854 SLSEMTFRPOLHSGDMVTPSPGLQLRLNEKLGTTADPLANDNHYGTQIPKEWKSQE 913  
QY 743 -----SONPPVLK 750  
Db 914 KSPEKTAFKKOTILSLNACSNHIAAINEGQNKPEIEVTWAKQGRTERLCSONPPVLK 973  
QY 751 RHOREITRITLQSQDEIDYDDTISVEMKEDFDIYDENENQSPRSQKTRHYFFIAAVE 810  
Db 974 RHOREITRITLQSQDEIDYDDTISVEMKEDFDIYDENENQSPRSQKTRHYFFIAAVE 1033  
QY 811 RLMDYGMSSSPHVLRLNRAQSGVPOFKKVVFOEFTDGSFTQPLYRGELNHLGLLGPYIR 870  
Db 1034 RLMDYGMSSSPHVLRLNRAQSGVPOFKKVVFOEFTDGSFTQPLYRGELNHLGLLGPYIR 1093  
QY 871 AEVEDNTMTVFRNQAQSPYFYSLSISYEEDQRCQABPRNFVKNPNETKTYFWKVQHMA 930  
Db 1094 AEVEDNTMTVFRNQAQSPYFYSLSISYEEDQRCQABPRNFVKNPNETKTYFWKVQHMA 1153  
QY 931 PTKDEPCKAWAYPSDVLKOVHSGLIGPLLVCHTNTLPAHGRQVTVQEFALFFTFID 990  
Db 1154 PTKDEPCKAWAYPSDVLKOVHSGLIGPLLVCHTNTLPAHGRQVTVQEFALFFTFID 1213  
QY 991 ETKSWYFTENMERNCRAPCNIQMEDPTFKENYRFAHNGYIMDTPLGLVMAQDQIRWYL 1050  
Db 1214 ETKSWYFTENMERNCRAPCNIQMEDPTFKENYRFAHNGYIMDTPLGLVMAQDQIRWYL 1273

QY 1051 LMSGNENIHSIHFGSHVFTVRKKEBYKALNLPFGVFTVEMLPKAGIWRVECLIGE 1110  
Db 1274 LMSGNENIHSIHFGSHVFTVRKKEBYKALNLPFGVFTVEMLPKAGIWRVECLIGE 1333  
QY 1111 HLHAGMSTLFLVYSNKKQCPPLGMAWHIRDFQITASQYGOVAPKLARLHYSGSINAWST 1170  
Db 1334 HLHAGMSTLFLVYSNKKQCPPLGMAWHIRDFQITASQYGOVAPKLARLHYSGSINAWST 1393  
QY 1171 KEPPFSWIKVLDLAPMIHGIKTQGARQKFFSSLYISQFIIMYSIDGKKWQTYRGNSTGTLM 1230  
Db 1394 KEPPFSWIKVLDLAPMIHGIKTQGARQKFFSSLYISQFIIMYSIDGKKWQTYRGNSTGTLM 1453  
QY 1231 VFPGNVDSGGIKHENIPNPIIARYIRLHPTHYSIRSLRMELMGCDLNSCMLPGMESKA 1290  
Db 1454 VFPGNVDSGGIKHENIPNPIIARYIRLHPTHYSIRSLRMELMGCDLNSCMLPGMESKA 1513  
QY 1291 ISDAQITASSYFTNMPATWSPSKARLHLOGRSNAMRPVNNPKBWLQVDFQKTMKVGT 1350  
Db 1514 ISDAQITASSYFTNMPATWSPSKARLHLOGRSNAMRPVNNPKBWLQVDFQKTMKVGT 1573  
QY 1351 TQGVKSLLTSMYKKEFLISSQDGHQWTLFFQNGKVKVFGQNDSTFPVNSLDPELLTR 1410  
Db 1574 TQGVKSLLTSMYKKEFLISSQDGHQWTLFFQNGKVKVFGQNDSTFPVNSLDPELLTR 1633  
QY 1411 YLRHQPQSWHQIALRMEVLGCEAQDLY 1438  
Db 1634 YLRHQPQSWHQIALRMEVLGCEAQDLY 1661

## RESULT 5

US-09-243-539-2  
; Sequence 2, Application US/09243539  
; Patent No. 6130203  
; GENERAL INFORMATION:  
; APPLICANT: VOORBERG, Johannes J.  
; TITLE OF INVENTION: HYBRID PROTEINS WITH MODIFIED ACTIVITY  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/243,539  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/558,107  
; FILING DATE: 13-NOV-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: ISACSON, John P.  
; REGISTRATION NUMBER: 33,715  
; REFERENCE/DOCKET NUMBER: 30472/212  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 672-5300  
; TELEFAX: (202) 672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1661 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-243-539-2

Query Match 96.78; Score 7434; DB 3; Length 1661;  
 Best Local Similarity 86.08; Pred. No. 0;  
 Matches 1416; Conservative 4; Mismatches 10; Indels 216; Gaps 4;

QY 1 ATRRYLGAVELSNDYMQSDGLBEPVDAAREPPRPVPSFPNTSVVYKTLFVEFTVHLEN 60  
 DB 20 ATRRYLGAVELSNDYMQSDGLBEPVDAAREPPRPVPSFPNTSVVYKTLFVEFTVHLEN 79  
 QY 61 IAKPRPMMGLGPTIOAEVYDVTVITLKNMASHPVSLHAGVSYKASEGAEYDDQTSQ 120  
 DB 80 IAKPRPMMGLGPTIOAEVYDVTVITLKNMASHPVSLHAGVSYKASEGAEYDDQTSQ 139  
 QY 121 REKEDDKVFPGGSHYVWVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 180  
 DB 140 REKEDDKVFPGGSHYVWVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 199  
 QY 181 EGSIAKKTOTLHKPILLFAVDFRGKSWHSETKNSLMQDDAASARAWPMHTVNGYVNR 240  
 DB 200 EGSIAKKTOTLHKPILLFAVDFRGKSWHSETKNSLMQDDAASARAWPMHTVNGYVNR 259  
 QY 241 SLPLGICHRKSVVWHVIGMTTPEVHSIFLEGHTFLVRNHRQASLEISPIFLTAQTL 300  
 DB 260 SLPLGICHRKSVVWHVIGMTTPEVHSIFLEGHTFLVRNHRQASLEISPIFLTAQTL 319  
 QY 301 MDLQOFLFCHISSHQHDGMEAYVKVDSCEPEPQLRMKNNEAEYDDDLTDSMDVVRP 360  
 DB 320 MDLQOFLFCHISSHQHDGMEAYVKVDSCEPEPQLRMKNNEAEYDDDLTDSMDVVRP 379  
 QY 361 DDNSPSFIQIRSAKHKPTWHYIAAEEDWDYAPLVLPDDRYSKYQYLNGPQRI 420  
 DB 380 DDNSPSFIQIRSAKHKPTWHYIAAEEDWDYAPLVLPDDRYSKYQYLNGPQRI 439  
 QY 421 RYKVKRFMAYTDTFTKTRAIQHESGILGLPLYGEVGDTLIIIFKNQASRPNIYPHGI 480  
 DB 440 RYKVKRFMAYTDTFTKTRAIQHESGILGLPLYGEVGDTLIIIFKNQASRPNIYPHGI 499  
 QY 481 TDVRPLYSRRLPKGVKHLKDFPILPGEIFPKYKMTVTVEDGPKSDPRCLTRYYSFVNME 540  
 DB 500 TDVRPLYSRRLPKGVKHLKDFPILPGEIFPKYKMTVTVEDGPKSDPRCLTRYYSFVNME 559  
 QY 541 RDLASGLIGPLLCYKESVDORGNQIMSDKRNVLFSVDENESWLTENIQRFLENPAG 600  
 DB 560 RDLASGLIGPLLCYKESVDORGNQIMSDKRNVLFSVDENESWLTENIQRFLENPAG 619  
 QY 601 VQLEDPEFQASNMHSGVVFQSLQSVCLHEVAYWYILSIGAQDTFLSVFSGYTFKH 660  
 DB 620 VQLEDPEFQASNMHSGVVFQSLQSVCLHEVAYWYILSIGAQDTFLSVFSGYTFKH 679  
 QY 661 KMVYEDTLTLPPSGEIVFMSMENPGLWILGCHNSDFRNGMTALLKVSSC----DKWTD 717  
 DB 680 KMVYEDTLTLPPSGEIVFMSMENPGLWILGCHNSDFRNGMTALLKVSSCIPGEBDD 739  
 QY 718 YY-----EDSYEDISAYLLKXNAIEPRSF-----742  
 DB 740 YLDLEKIFSEDDYIDI-----VDSLEPRFSQNSRHPSTKQFNATTIPENDIEKT 793  
 QY 743 -----742  
 DB 794 DPMFAHRTMPKIQNVSSDLLMLLRQSPTPHGLSLASDLQEAKEYTFSDPSPGAIDSN 853  
 QY 743 -----742  
 DB 854 SLSEWTHFRPOLHSGDNWFTPSGLQRLNEKLGTTADPLAWNHYGQIPKEWKSOR 913  
 QY 743 -----742  
 DB 914 KSPEKTAFFKXDTILSNACSNHIAIAINEGQNKPEIEVYMAKQGRTERLCSQNPVLK 973  
 QY 751 RHOREITRTTLOSQOEIDYDDTISVEMKEDEDIYDENQSPRFSQKTRHYFTAAVE 810  
 DB 974 RHOREITRTTLOSQOEIDYDDTISVEMKEDEDIYDENQSPRFSQKTRHYFTAAVE 1033  
 QY 811 RLWDYGNSSSPHVLNRNAQSGSVPOPKVQVQFETDGSFTQPLYRGELNEHLGLGPYIR 870

DB 1034 RLWDYGNSSSPHVLNRNAQSGSVPOPKVQVQFETDGSFTQPLYRGELNEHLGLGPYIR 1093  
 QY 871 AEVEDNIMVTRNOASRPYSFYSLISYEBDQRCQABPRKNFKVKNPNETKTYFKVQHHMA 930  
 DB 1094 AEVEDNIMVTRNOASRPYSFYSLISYEBDQRCQABPRKNFKVKNPNETKTYFKVQHHMA 1153  
 QY 931 PTKDEFCKAWAYESDVLEKDVHSGGLIGPLLVCHTNTLPAHGRQVTVQEPALFTI 990  
 DB 1154 PTKDEFCKAWAYESDVLEKDVHSGGLIGPLLVCHTNTLPAHGRQVTVQEPALFTI 1213  
 QY 991 ETKSWYFTENMERNCRAPCIQMEDPTFKENYRFAHNGIYMDTLPGLVNAQOQIRWYL 1050  
 DB 1214 ETKSWYFTENMERNCRAPCIQMEDPTFKENYRFAHNGIYMDTLPGLVNAQOQIRWYL 1273  
 QY 1051 LSMGSNENIHSIHESGHVFTVRKKEEYKMAIYNYLPGVFTVEMLPKAGIWRVECLIGE 1110  
 DB 1274 LSMGSNENIHSIHESGHVFTVRKKEEYKMAIYNYLPGVFTVEMLPKAGIWRVECLIGE 1333  
 QY 1111 HLHAGMSTLFLVYSNKQCTPLGMAHGHIRDFQITASQYQWAPKLARLHYSGSINAWST 1170  
 DB 1334 HLHAGMSTLFLVYSNKQCTPLGMAHGHIRDFQITASQYQWAPKLARLHYSGSINAWST 1393  
 QY 1171 KEPPSWIKVDLLAPMIIHGINTQCARQKFSLSYISQFIIMYSLDGKKWQTYRGNSTGTLM 1230  
 DB 1394 KEPPSWIKVDLLAPMIIHGINTQCARQKFSLSYISQFIIMYSLDGKKWQTYRGNSTGTLM 1453  
 QY 1231 VFGNVDSSGGLKHNIFNPPIIARIYLRHPTHYSTRLMELMGLCDLNSCMLPGMESKA 1290  
 DB 1454 VFGNVDSSGGLKHNIFNPPIIARIYLRHPTHYSTRLMELMGLCDLNSCMLPGMESKA 1513  
 QY 1291 ISDAQITASSYFTNMFTWSPSKARLHQGRSNARVQVNNPKWQLQVDFQTKMKVTGVT 1350  
 DB 1514 ISDAQITASSYFTNMFTWSPSKARLHQGRSNARVQVNNPKWQLQVDFQTKMKVTGVT 1573  
 QY 1351 TQGVKSLTSMYVKEFLISSQDGHQWTLFPQNGKVKVFOGNOQSDFTPVVNSLDPPLLTR 1410  
 DB 1574 TQGVKSLTSMYVKEFLISSQDGHQWTLFPQNGKVKVFOGNOQSDFTPVVNSLDPPLLTR 1633  
 QY 1411 YLRHPOSWVHQAIALRMEVLGCEADLY 1438  
 DB 1634 YLRHPOSWVHQAIALRMEVLGCEADLY 1661

RESULT 6  
 US-07-864-004B-4  
 ; Sequence 4, Application US/07864004B  
 ; Patent No. 5364771  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lollar, John S.  
 ; APPLICANT: Runge, Marshall S.  
 ; TITLE OF INVENTION: Hybrid Human/Porcine Factor VIII  
 ; NUMBER OF SEQUENCES: 6  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Kilpatrick & Cody  
 ; STREET: 1100 Peachtree Street  
 ; CITY: Atlanta  
 ; STATE: Georgia  
 ; COUNTRY: US  
 ; ZIP: 30309  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent in Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/864,004B  
 ; FILING DATE: 07 APRIL 1992  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Pabst, Patrea L.  
 ; REGISTRATION NUMBER: 31,284  
 ; REFERENCE/DOCKET NUMBER: EMU106

## TELECOMMUNICATION INFORMATION:

TELEPHONE: 404-815-6508

TELEFAX: 404-815-6555

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 2332 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ANTI-SENSE: NO N-terminal

FRAGMENT TYPE: NO

ORGANISM: Homo sapien

ORIGINAL SOURCE:

TISSUE TYPE: Liver cDNA sequence

US-07-864-004B-4

Query Match 94.1%; Score 7234; DB 1; Length 2332;

Best Local Similarity 61.7%; Pred. No. 0;

Matches 1438; Conservative 0; Mismatches 0; Indels 894; Gaps 1;

QY 1 ATRPVYLGAVELSDWYQSDGLPVDAREPPVPKSPFPENTSVYKKTLLFVEFTVHLFN 60  
 DB 1 ATRPVYLGAVELSDWYQSDGLPVDAREPPVPKSPFPENTSVYKKTLLFVEFTVHLFN 60  
 QY 61 IAKPRPPWMLLGTPTQAEYVDVITLKNMASHPVSLHAGVSYWKASGEAYDDQTSQ 120  
 DB 61 IAKPRPPWMLLGTPTQAEYVDVITLKNMASHPVSLHAGVSYWKASGEAYDDQTSQ 120  
 QY 121 REKEDKVPFGSGHTYVQVLKENGPMASDPLCLITYSLSHVDLVKDLNSGLIGALLVCR 180  
 DB 121 REKEDKVPFGSGHTYVQVLKENGPMASDPLCLITYSLSHVDLVKDLNSGLIGALLVCR 180  
 QY 181 EGS LAKEKOTLHKFILLFAVDEGKSWHSETKNSLMQDRDAASARAWPKGHTVNGVNR 240  
 DB 181 EGS LAKEKOTLHKFILLFAVDEGKSWHSETKNSLMQDRDAASARAWPKGHTVNGVNR 240  
 QY 241 SLPLGIGCHRKSVYHVGITGTPPEVHSIFLEGHTFLVRNHRQASLEISPIITLTAQTLL 300  
 DB 241 SLPLGIGCHRKSVYHVGITGTPPEVHSIFLEGHTFLVRNHRQASLEISPIITLTAQTLL 300  
 QY 301 MDLGQFLLPCHISSHQHDGMEAYVKVDCSPEEPQLRMKNNEAEYDDDLTDSEMDVVRP 360  
 DB 301 MDLGQFLLPCHISSHQHDGMEAYVKVDCSPEEPQLRMKNNEAEYDDDLTDSEMDVVRP 360  
 QY 361 DDNSPSFIQIRSVAKKPKTWVHYIAAREEDWDYAPLVLAPODRSVKSOYIANGPQRI 420  
 DB 361 DDNSPSFIQIRSVAKKPKTWVHYIAAREEDWDYAPLVLAPODRSVKSOYIANGPQRI 420  
 QY 421 RYKVKVRFMAYTDETFKTREAIQHSIGILGPLLYGEVGTLLIIFKNQASRPNIYPHGI 480  
 DB 421 RYKVKVRFMAYTDETFKTREAIQHSIGILGPLLYGEVGTLLIIFKNQASRPNIYPHGI 480  
 QY 481 TDVRLYSRRLPKGVKHLKDPFIPILPGEIFKYKWTVTVEDGPTKSDPCLFRYYSFVNME 540  
 DB 481 TDVRLYSRRLPKGVKHLKDPFIPILPGEIFKYKWTVTVEDGPTKSDPCLFRYYSFVNME 540  
 QY 541 RDLASGLIGPLLIICYKESVDQNGQIMSDGNVILFSVPDENRHWLTENIQRLPNPAG 600  
 DB 541 RDLASGLIGPLLIICYKESVDQNGQIMSDGNVILFSVPDENRHWLTENIQRLPNPAG 600  
 QY 601 VQLDEPFOASNIWHSINGVVFOSLQSVCLHEVAYWYIISGAQTDPLSVFSPSGYTFKH 660  
 DB 601 VQLDEPFOASNIWHSINGVVFOSLQSVCLHEVAYWYIISGAQTDPLSVFSPSGYTFKH 660  
 QY 661 KMVYEDTLTLFPFSGETVFMASMPGLWILGCHNSDFRNRGMTALLKVSSCDKNTGDYYE 720  
 DB 661 KMVYEDTLTLFPFSGETVFMASMPGLWILGCHNSDFRNRGMTALLKVSSCDKNTGDYYE 720  
 QY 721 DSVEDISAYLLSKNNAIEPRSF----- 742  
 DB 721 DSVEDISAYLLSKNNAIEPRSF----- 742

QY 743 ----- 742  
 DB 781 KIQVSSDILLRLRQSPFPHGLSLSDLOEAKYETFDSPDFCAIDSNNSLSEMTHERPQ 840  
 QY 743 ----- 742  
 DB 841 LHHSGDMVTPESGLQLRLNEKLGTAAATELKKLDKFKVSSTSNLLISTIPSNLAAGTON 900  
 QY 743 ----- 742  
 DB 901 TSSLGPPSPMPHYDVSQDITTLFGKKSSPLTESGGPLSLEKNDSKLLSGLMWSQESSW 960  
 QY 743 ----- 742  
 DB 961 GKNVSTESGRLFKGRKRAHPALLTKDNALFKVISISLLATNKTNSNATWRTKTHIDGPSL 1020  
 QY 743 ----- 742  
 DB 1021 LIENSPSVWONILESTEPKVTPLIHDRMLMDKNATALLRLHMSNKTTSKNMVMVQOK 1080  
 QY 743 ----- 742  
 DB 1081 KEGPIPPDAQNPDMSFFKMLFLPESARWLTQRTGKNSLNSGGSPKQLVSLGPEKSVEG 1140  
 QY 743 ----- 742  
 DB 1141 QNFLEKKNVVVGKGEFTKDVGLKEMVPPSSRNFLTNLNLTNNHNNTHNQEKIOBEIEK 1200  
 QY 743 ----- 742  
 DB 1201 KETLIOENVLPQIHTVTGTKNFMKNLFLSTRQNVESYEGAYAPVLQDFRSLNDSTNR 1260  
 QY 743 ----- 742  
 DB 1261 TKGHTAHFSPKKEENLEGLGNQTKQIYEKYACTTISPTNSQONFVQTSKALQFRL 1320  
 QY 743 ----- 742  
 DB 1321 PLEETELEKRIIVDDTSTQSKNMKHLPTSLTQIDYNEKEKGAITQSPSLDCLTRSHSI 1380  
 QY 743 ----- 742  
 DB 1381 PQANRSPPLIAKVSSFPPIRPIYLRVLPQDNSSHLPAASYRKDSGVQESSHFLAGAK 1440  
 QY 743 ----- 742  
 DB 1441 NNLSLAILLEMTGDOREVGSLGTSATNSVYKKVENTVLPKPDLPKTSKVELLPKVHI 1500  
 QY 743 ----- 742  
 DB 1501 YQKDLFPPTETSNPGHLDLVEGSLQGTGEGAIAKNNEANRPGKVPFLRVATESSAKTPSK 1560  
 QY 743 ----- 742  
 DB 1561 LLDPLAWNHYGTQIPKEBWKSOEKSPEKTAFFKKDPTILSLANACESNHAIAAINQONKP 1620  
 QY 743 -----SONPPVLKRHOREITRTLQSDQDEEIDYDDTISVEMKKEBFDIY 786  
 DB 1621 EIEVTWAKQGRTERLCSQNPVPLKRHOREITRTLQSDQDEEIDYDDTISVEMKKEBFDIY 1680  
 QY 787 DEDENQSPRSFQKTKRHYFIAAVERLDWYGMSSSPHVLNRNAQSGSVQPKKVVQFEFTD 846  
 DB 1681 DEDENQSPRSFQKTKRHYFIAAVERLDWYGMSSSPHVLNRNAQSGSVQPKKVVQFEFTD 1740  
 QY 847 GSTQPLYRGELNEHGLLGPYIRAEVDENIMVTFRNQASRPYSFYSLLSISEEDORQGA 906  
 DB 1741 GSTQPLYRGELNEHGLLGPYIRAEVDENIMVTFRNQASRPYSFYSLLSISEEDORQGA 1800  
 QY 907 EPEKNFVKNETKTYFWKQVQHMAPTKDFDCKAMAYPSDVDLEKDVHSGSLIGLPLVCHT 966  
 DB 1801 EPRKNFVKNETKTYFWKQVQHMAPTKDFDCKAMAYPSDVDLEKDVHSGSLIGLPLVCHT 1860



967 NTLPAGHROVTVQEFALFFTFIDETKSWYFTENMERNCRAPCNIMEDPTFKENYRPHA 1026  
1861 NTLPAGHROVTVQEFALFFTFIDETKSWYFTENMERNCRAPCNIMEDPTFKENYRPHA 1920  
1027 INGIMDTLPLGVAQDORIRWYLLSGMSNENIHSIHPSGHVFTVRKKEEKMALYNLYP 1086  
1921 INGIMDTLPLGVAQDORIRWYLLSGMSNENIHSIHPSGHVFTVRKKEEKMALYNLYP 1980  
1087 GVFTVEMLPKAGIRWVECLIGELHAGMSTLFLVYSNKCOTPLGMAHGHIRDFQITAS 1146  
1981 GVFTVEMLPKAGIRWVECLIGELHAGMSTLFLVYSNKCOTPLGMAHGHIRDFQITAS 2040  
1147 GOYGQWAPKLARLHYSGINAWSTKEPFSWIKVDLLAPMIHGIKTQGARQKFSLSYISQ 1206  
2041 GOYGQWAPKLARLHYSGINAWSTKEPFSWIKVDLLAPMIHGIKTQGARQKFSLSYISQ 2100  
1207 FIIMYSLDGKKQWTVRGNGSTGTLMVFFGNVDSSGIGKNI FNPPPIARIYIRLHPHYSIRS 1266  
2101 FIIMYSLDGKKQWTVRGNGSTGTLMVFFGNVDSSGIGKNI FNPPPIARIYIRLHPHYSIRS 2160  
1267 TLRVLMGCDLNSCMPJGMSKALSDAQITASSYFTNMFTWSPSKARLHLQGRSNAWR 1326  
2161 TLRVLMGCDLNSCMPJGMSKALSDAQITASSYFTNMFTWSPSKARLHLQGRSNAWR 2220  
1327 PQVNNPKWLQVDFOKTMKVTVGTTQGVKSLTSMYKFEFLISSQDGHQWTLFPQNGKV 1386  
2221 PQVNNPKWLQVDFOKTMKVTVGTTQGVKSLTSMYKFEFLISSQDGHQWTLFPQNGKV 2280  
1387 KVFQGNQSFPTVNSLDPLLTRYLRIHPQSWHQAIALRMEVLGCEAQDLY 1438  
2281 KVFQGNQSFPTVNSLDPLLTRYLRIHPQSWHQAIALRMEVLGCEAQDLY 2332

RESULT 7

US-08-251-937A-4  
Sequence 4, Application US/08251937A  
Patent No. 5583209  
GENERAL INFORMATION:  
APPLICANT: Lollar, John S.  
APPLICANT: Runge, Marshall S.  
TITLE OF INVENTION: Hybrid Human/Porcine Factor VIII  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Kilpatrick & Cody  
STREET: 1100 Peachtree Street  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: US  
ZIP: 30309  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/251,937A  
FILING DATE: 31-MAY-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/864,004  
FILING DATE: 07-APR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Pratt, John S.  
REGISTRATION NUMBER: 29,476  
REFERENCE/DOCKET NUMBER: EMU106DIV  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-815-6367  
TELEFAX: 404-815-6555  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2332 amino acids  
TYPE: amino acid  
STRANDEDNESS: single

TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Homo sapien  
TISSUE TYPE: Liver cdna sequence  
US-08-251-937A-4  
Query Match 94.1%; Score 7234; DB 1; Length 2332;  
Best Local Similarity 61.7%; Pred. No. 0;  
Matches 1438; Conservative 0; Mismatches 0; Indels 894; Gaps 1;  
QY 1 ATRYIYLGAVELSDWYMQSDLGELPVDARPPRPVKSPFPNTSVYVYKTLFVEFTVHLFN 60  
DB 1 ATRYIYLGAVELSDWYMQSDLGELPVDARPPRPVKSPFPNTSVYVYKTLFVEFTVHLFN 60  
QY 61 IAKRPPPMGLLGFTIOAEVYDVTVIITLKNMASHPVSLHAGVSYKASGAEYDDQTSQ 120  
DB 61 IAKRPPPMGLLGFTIOAEVYDVTVIITLKNMASHPVSLHAGVSYKASGAEYDDQTSQ 120  
QY 121 REKEDDKVFPGGSHYVQVLEKNGPMASDPLCLITYSLSHVDLVKDINSGLIGALLVCR 180  
DB 121 REKEDDKVFPGGSHYVQVLEKNGPMASDPLCLITYSLSHVDLVKDINSGLIGALLVCR 180  
QY 181 EGSIAKKTQTLHKPFIILLFAVFDGKSWHSTKNSLMODRDAASARAWPKMHTVNGYVNR 240  
DB 181 EGSIAKKTQTLHKPFIILLFAVFDGKSWHSTKNSLMODRDAASARAWPKMHTVNGYVNR 240  
QY 241 SLPLGIGCHRSVYVHVGTTTPEVHSIFLEGHTFLVRNHRQASLEISPIITFLTAQLL 300  
DB 241 SLPLGIGCHRSVYVHVGTTTPEVHSIFLEGHTFLVRNHRQASLEISPIITFLTAQLL 300  
QY 301 MDLQGFLLFCHISSHQHDGMEAYVKVDSCEPQLRMKNNEAEYDDDLTDSEMDVYVRF 360  
DB 301 MDLQGFLLFCHISSHQHDGMEAYVKVDSCEPQLRMKNNEAEYDDDLTDSEMDVYVRF 360  
QY 361 DDNSPSFIQIRSVAKKHPTWVHYIAAEEEDWDYAPLVLPADDDRSYKQYLNNQPORG 420  
DB 361 DDNSPSFIQIRSVAKKHPTWVHYIAAEEEDWDYAPLVLPADDDRSYKQYLNNQPORG 420  
QY 421 RYKVKRPMAYTDETFKTREAIQHSGLIGLPLLYGEVGDTLIIIFKNQASRYNLYPHGI 480  
DB 421 RYKVKRPMAYTDETFKTREAIQHSGLIGLPLLYGEVGDTLIIIFKNQASRYNLYPHGI 480  
QY 481 TDVRLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGPTKSDPRCLTRYSSFVNM 540  
DB 481 TDVRLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGPTKSDPRCLTRYSSFVNM 540  
QY 541 RDLASGLIGPLLI CYKESVDQRGNQIMSDKRNVLIFSVPDENRSWYLTENIORFLPNPAG 600  
DB 541 RDLASGLIGPLLI CYKESVDQRGNQIMSDKRNVLIFSVPDENRSWYLTENIORFLPNPAG 600  
QY 601 VQLEDPEQASNIMHSINGYVDSIQLSVCLHEVAYWYTLISGAOTDFLSVFFSGYTFKH 660  
DB 601 VQLEDPEQASNIMHSINGYVDSIQLSVCLHEVAYWYTLISGAOTDFLSVFFSGYTFKH 660  
QY 661 KMVYEDTLTLFPFSGETVFMSENGLITLGNCHSDFRNRGMGTALLIKVSSCDKNTGDYYE 720  
DB 661 KMVYEDTLTLFPFSGETVFMSENGLITLGNCHSDFRNRGMGTALLIKVSSCDKNTGDYYE 720  
QY 721 DSYEDISAYLLSKNNAIEPRSF----- 742  
DB 721 DSYEDISAYLLSKNNAIEPRSFQNSRHPSTRQKQFNATTIPENDIEKTDPMFAHRTMP 780  
QY 743 ----- 742  
DB 781 KI QNVSSDILLMLLRQSPTPHGLSLDLOEAKYETFSDDPSGADSNNSLSEMTFRPQ 840  
QY 743 ----- 742  
DB 841 LHHSGDMVTPSPESGLQRLNEKLGTTAATLKKLDFKVSSTSNLLISTIPSDNLAAGTDN 900

QY 743 ----- 742  
Db 901 TSSLGPPMPVHYDSQDITTLFGKSSPLTESGGPLSSENNDSKLESGLMNSQESSW 960  
QY 743 ----- 742  
Db 961 GKNVSTESGRLPFGKRAHPALLTKONALFKVSIISLLTKNTKSNNSATNRKTHIDGPSL 1020  
QY 743 ----- 742  
Db 1021 LIENSPVQNIILESTEFKVTPLIHDRMLMDKNATALRLNHSNKTSSKNWVQOK 1080  
QY 743 ----- 742  
Db 1081 KEGPIPPDAQPDMSPFKMLFLPESARWIQRTHGKNSLNSGGQSPKQLVSLGPEKSVEG 1140  
QY 743 ----- 742  
Db 1141 QNFLSEKNVVGKGEFTKDVGLKEMVFPSSRNLFTNLNLHNNTHNQEKKIQEBIEK 1200  
QY 743 ----- 742  
Db 1201 KETLIQENVVLPOIHVTGTGNFKMKNLFLLSLSTRQNVESYEGAYAPVLQDFRSINDSTNR 1260  
QY 743 ----- 742  
Db 1261 TKKHTAHFSKGBEENLEGLGNQTKOIVEKYACTTRISPNTSQNFVQORSKRALKQFRL 1320  
QY 743 ----- 742  
Db 1321 PLEETELEKRIIIVDDTSTOWSKNMKHLTPSLTAQIDYNEKEGAIQTSPSLDCLTRSHSI 1380  
QY 743 ----- 742  
Db 1381 PQANRSPLPKAVSSPFSIRPIVLRVLFQDNSSHLPAASYRKDSGVQESSHFLQAGK 1440  
QY 743 ----- 742  
Db 1441 NNLSLAILTLEMTGDQREVSLGTSATNSVYKKVENTVLPKPDLPKTSKGVKELLPKVHI 1500  
QY 743 ----- 742  
Db 1501 YQKDLPTTETNSGPHLDIVESLLQGTGKALWNEANRPGKVPFLRVATESAKTPSK 1560  
QY 743 ----- 742  
Db 1561 LLDPLAWNHYGTQIPKEEWKSEKSPKTAFFKKDTILSLNACSNHATAAINNEGONKP 1620  
QY 743 -----SQNPVLRKHOREITRITLQSDQBEIDYDDTISVEMKEDFDIY 786  
Db 1621 EIEVTWAKQGRTERLCSQNPVLRKHOREITRITLQSDQBEIDYDDTISVEMKEDFDIY 1680  
QY 787 DEENOSPRSFOKKTTHYFTAAVERLWDYQWSSSPHVLNRNRAQSGVPQKKVVFQBFETD 846  
Db 1681 DEENOSPRSFOKKTTHYFTAAVERLWDYQWSSSPHVLNRNRAQSGVPQKKVVFQBFETD 1740  
QY 847 GSFTQPLYRGELNEHLGLGPYRAEVEDNIMWTFRNOASRPYSFYSSLLISYVEDORQGA 906  
Db 1741 GSFTQPLYRGELNEHLGLGPYRAEVEDNIMWTFRNOASRPYSFYSSLLISYVEDORQGA 1800  
QY 907 EPRKNFKVNETKTYFWKQHRHAPTKDEFDCAKAWYFSDVLEKDVHSGILGLPLVCHT 966  
Db 1801 EPRKNFKVNETKTYFWKQHRHAPTKDEFDCAKAWYFSDVLEKDVHSGILGLPLVCHT 1860  
QY 967 NTLNPAHGRQVTVQEFALFTTIFDETSKYFTENMERNCRAPCNQIMEDPTFKENYRPHA 1026  
Db 1861 NTLNPAHGRQVTVQEFALFTTIFDETSKYFTENMERNCRAPCNQIMEDPTFKENYRPHA 1920  
QY 1027 INGYIMDTLPLVMAQDQRIWVLLSMGSENTHSFSGHVTFRKKEBKALYNLYP 1086  
Db 1921 INGYIMDTLPLVMAQDQRIWVLLSMGSENTHSFSGHVTFRKKEBKALYNLYP 1980

QY 1087 GYFETVEMLPKAGIWRVECLIGEHLHAGMSTFLVYSNKCOTPLGMASGHIRDFOITAS 1146  
Db 1981 GYFETVEMLPKAGIWRVECLIGEHLHAGMSTFLVYSNKCOTPLGMASGHIRDFOITAS 2040  
QY 1147 GQYGWAPKARLHYSGSINAMSTKEPFSWIKVDLLAPMIHGIKTQGARQKFSSLYISQ 1206  
Db 2041 GQYGWAPKARLHYSGSINAMSTKEPFSWIKVDLLAPMIHGIKTQGARQKFSSLYISQ 2100  
QY 1207 FIIMYSLDGKKQVTRGNSGTGLMVFVFGNVDSGKGNHFNPPITARYIRLHPTHYSIRS 1266  
Db 2101 FIIMYSLDGKKQVTRGNSGTGLMVFVFGNVDSGKGNHFNPPITARYIRLHPTHYSIRS 2160  
QY 1267 TLRMELMGDLNSCSMPLGMEKASIDAOITASSYFTNMFATWSPSKARLHLQGRSNWR 1326  
Db 2161 TLRMELMGDLNSCSMPLGMEKASIDAOITASSYFTNMFATWSPSKARLHLQGRSNWR 2220  
QY 1327 PQVNNPKWLQVDFOKTMKVTVGTQGVKSLLTSMYVKEFLISSSDGHWTLFFQNGKV 1386  
Db 2221 PQVNNPKWLQVDFOKTMKVTVGTQGVKSLLTSMYVKEFLISSSDGHWTLFFQNGKV 2280  
QY 1387 KVFQGNQDSFTPVNSLDPPLLTRYLRIHPQSVWVHQIALRMEVLGCEADLY 1438  
Db 2281 KVFQGNQDSFTPVNSLDPPLLTRYLRIHPQSVWVHQIALRMEVLGCEADLY 2332

## RESULT 8

US-08-212-133A-2  
; Sequence 2, Application US/08212133A  
; Patent No. 5663060  
; GENERAL INFORMATION:  
; APPLICANT: Lollar, John S.  
; APPLICANT: Runge, Marshall S.  
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kilpatrick & Cody  
; STREET: 100 Peachtree Street  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: US  
; ZIP: 30303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/212,133A  
; FILING DATE: March 11, 1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/864,004  
; FILING DATE: 07-APR-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Padst, Patrea L.  
; REGISTRATION NUMBER: 31,284  
; REFERENCE/DOCKET NUMBER: EMU/76677  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 404-572-6508  
; TELEFAX: 404-572-6555  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2332 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapien  
; TISSUE TYPE: Liver cDNA sequence

US-08-212-133A-2

Query Match		94.1%;	Score 7234;	DB 1;	Length 2332;
Best Local Similarity		61.7%;	Pred. No. 0;		
Matches 1438; Conservative		0;	Mismatches	0;	Indels 894; Gaps 1;
QY	1	ATRYILGAVELSDYMQSDIGELPVDARPPRPVKSPFPNTSVVYKTLFVEFTVHLFN	60		
DB	1	ATRYILGAVELSDYMQSDIGELPVDARPPRPVKSPFPNTSVVYKTLFVEFTVHLFN	60		
QY	61	IAPRRPMMGLLGTPTIAQEVYDVTIVITLKNMASHPVSLHVGVSYNKASGAEVDDQTSQ	120		
DB	61	IAPRRPMMGLLGTPTIAQEVYDVTIVITLKNMASHPVSLHVGVSYNKASGAEVDDQTSQ	120		
QY	121	REKEDDKVPFGGSHYVQVILKENGPMASDPLCLITYSLSHVDLVKDLNSGLICALLVCR	180		
DB	121	REKEDDKVPFGGSHYVQVILKENGPMASDPLCLITYSLSHVDLVKDLNSGLICALLVCR	180		
QY	181	EGSLAKETQTLHKFILLFAVFDGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNR	240		
DB	181	EGSLAKETQTLHKFILLFAVFDGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNR	240		
QY	241	SLPGLIGCHRKSVMVHVGMTTPEVHSIFLEGHTFLVRNROASLEISPIITFLTAQTL	300		
DB	241	SLPGLIGCHRKSVMVHVGMTTPEVHSIFLEGHTFLVRNROASLEISPIITFLTAQTL	300		
QY	301	MDLQGFLLFCHISSHQDHGMEAYVKVDSCPEEPQLRMKNNEAEYDDDLTDSEMDVVRP	360		
DB	301	MDLQGFLLFCHISSHQDHGMEAYVKVDSCPEEPQLRMKNNEAEYDDDLTDSEMDVVRP	360		
QY	361	DDNSPSPQIRSVAKHPTKTVHYIAAEEDWDYAPLVLPAPDDRYSKQYLNNPQIRG	420		
DB	361	DDNSPSPQIRSVAKHPTKTVHYIAAEEDWDYAPLVLPAPDDRYSKQYLNNPQIRG	420		
QY	421	RKYKVRMAYTDETFKTRALQHSIGLPLLYGEVGDTLIIIFKNOASRPYNIYPHGI	480		
DB	421	RKYKVRMAYTDETFKTRALQHSIGLPLLYGEVGDTLIIIFKNOASRPYNIYPHGI	480		
QY	481	TDVRLYSRRLPKGVKHLKDPILPGEIFKYKWTVVEDGPTKSDPRCLTRYSSFVNME	540		
DB	481	TDVRLYSRRLPKGVKHLKDPILPGEIFKYKWTVVEDGPTKSDPRCLTRYSSFVNME	540		
QY	541	RDASGLIGPLLI CYKESVDQGNQIMSDKRNVLIFSVDENRSLYLTENIORFLPNPAG	600		
DB	541	RDASGLIGPLLI CYKESVDQGNQIMSDKRNVLIFSVDENRSLYLTENIORFLPNPAG	600		
QY	601	VOLEDPFOASNMHSINGYVDFSLSVCLHEVAYWYILSGAOTDFLSVFFSGVTFKH	660		
DB	601	VOLEDPFOASNMHSINGYVDFSLSVCLHEVAYWYILSGAOTDFLSVFFSGVTFKH	660		
QY	661	KMYVEDTLTLPFPGSTVPMSENPGMLWILGCHNSDFRNKGMTALIKVSSCDKNTGDYYE	720		
DB	661	KMYVEDTLTLPFPGSTVPMSENPGMLWILGCHNSDFRNKGMTALIKVSSCDKNTGDYYE	720		
QY	721	DSYEDISAYLLSKNNAIEPRSF	742		
DB	721	DSYEDISAYLLSKNNAIEPRSFQNSRHPSTROKQFNATTIPENDIEKTDPMFAHRTMP	780		
QY	743	-----	742		
DB	781	KIONVSSDILLMLLRQSPHGLSLDQAEKYTEFDDPPSGAIDSNNSLSEMTFRPQ	840		
QY	743	-----	742		
DB	841	LHSGDMVTPPESGLQRLNEKLGTTAATLKKLDPKVSTSNLITIPSDNLAAGTDN	900		
QY	743	-----	742		
DB	901	TSSIGPSPMPVHYDSQDITTLFGKSSPLTESGGPLSLSENNDSKLLSGLMNSQSSW	960		
QY	743	-----	742		
DB	961	GKNVSTESGRLFKGRAHPALLTKDNALFKVSIILLKTNKTSNNSATNRKTHIDGPSL	1020		

QY	743	-----	742
DB	1021	LIENSPVMQNILETSETEFKVTPLIHDRMLMDKNATALLRLNMSNKTTSKKNMEMVQOK	1080
QY	743	-----	742
DB	1081	KEGPIPPDAQNDPMSFFKMLFLPESARWIORTHGKNLSNGQGPSKQLVSLGPEKSVEG	1140
QY	743	-----	742
DB	1141	QNFLSEKNVVGKGEFTKDVGLKEMVFPSSRNFLTLNLDNLHNNTHNQEKIOBEIEK	1200
QY	743	-----	742
DB	1201	KETLIOENVVLPOIHVTGTGKNFMKNLFLSTRQNVESGYEGAYAPVLQFRSLNSTNR	1260
QY	743	-----	742
DB	1261	TKKHTAHFSKKGEHENLEGLGNQTKQIVEKVACTTRISPNTSQNFVTOQRSKRALKQFRL	1320
QY	743	-----	742
DB	1321	PLBETELEKRIIVDDTSTQSKNMKHLTPSTLTQIDYNEKEKGAITQSPSLDCLTRSHSI	1380
QY	743	-----	742
DB	1381	POANRSPPIAKVSSFPSPSIRPIYLRVLFDNSSHLPAASTRYKDKSGVQSSSHFLOGAKK	1440
QY	743	-----	742
DB	1441	NNLSLAITLMTQDQREVGSLGTSATNSVTKKVENTVLPKPDLPKTSKGVKVELLPKVIH	1500
QY	743	-----	742
DB	1501	YQDLFPETTSNGSPGHLDLVEGSLLOQTEGAIKWEANRPGKVPFLRVATESSAKTPSK	1560
QY	743	-----	742
DB	1561	LLDPLANDNHVGTQIPKEWKSQKSPKTAFFKDDTILSNACESNHAIAINEGQNKP	1620
QY	743	-----SONPPVLKRHQREIIRTTLSQDQEEIDYDDTISVEMKKEDFDIY	786
DB	1621	EIFVTWAKQGTERTLCSONPPVLKRHQREIIRTTLSQDQEEIDYDDTISVEMKKEDFDIY	1680
QY	787	DEDENQSPRSQKTRHYFIAAVERLWDYGMSSSHVLNRNAQSGSVQFKKVFQFETD	846
DB	1681	DEDENQSPRSQKTRHYFIAAVERLWDYGMSSSHVLNRNAQSGSVQFKKVFQFETD	1740
QY	847	GSFTQPLRGELNEHGLGPGYIRAEVEDNIMVTFRNQASRPYSFVSSLSIYEEEDQROGA	906
DB	1741	GSFTQPLRGELNEHGLGPGYIRAEVEDNIMVTFRNQASRPYSFVSSLSIYEEEDQROGA	1800
QY	907	EPKKNFVKNPNETKYFWKVQHMAPTKDFDCKAWAYFSDVDLEKDVHSGGLIGPLLVCHT	966
DB	1801	EPKKNFVKNPNETKYFWKVQHMAPTKDFDCKAWAYFSDVDLEKDVHSGGLIGPLLVCHT	1860
QY	967	NTLNPAGHQVTVQFALFFTFIDETKSNWYFTENMERNCRAPCNTOMEDPTFKENYRPHA	1026
DB	1861	NTLNPAGHQVTVQFALFFTFIDETKSNWYFTENMERNCRAPCNTOMEDPTFKENYRPHA	1920
QY	1027	INGYIMDTLPGLVMAQDQRIRWYLLSMGSNENIHSIFSGHVFTVRKKEEYKWAYLYNP	1086
DB	1921	INGYIMDTLPGLVMAQDQRIRWYLLSMGSNENIHSIFSGHVFTVRKKEEYKWAYLYNP	1980
QY	1087	GYFETVEMLPKSKAGIWRVECLIGEHLHAGMSTFLVYSNKCOTPLGMSGHIRDQITAS	1146
DB	1981	GYFETVEMLPKSKAGIWRVECLIGEHLHAGMSTFLVYSNKCOTPLGMSGHIRDQITAS	2040
QY	1147	GOYGOWAPKLARLHSGSINAWSTKEPFSWIKVDLLAPMIHGIKTQAGARQFSSLIYSQ	1206
DB	2041	GOYGOWAPKLARLHSGSINAWSTKEPFSWIKVDLLAPMIHGIKTQAGARQFSSLIYSQ	2100

QY 1207 FIIMSLDGKKWOTYRGNSTGTLWVFGNVDSSGIKININFPPIIARIIRLHPTHYSIRS 1266  
DB 2101 FIIMSLDGKKWOTYRGNSTGTLWVFGNVDSSGIKININFPPIIARIIRLHPTHYSIRS 2160  
QY 1267 TLRLMELMGCDLNSCMLGMSKALISDAQITASSYFTNMFATWSPSKARLHLQGRSNWR 1326  
DB 2161 TLRLMELMGCDLNSCMLGMSKALISDAQITASSYFTNMFATWSPSKARLHLQGRSNWR 2220  
QY 1327 PQVNNPKEMQLQVDFQKTMKVTVGTTQGVKSLTSMYKKEFLISSQDGHQWTLFFQNGKV 1386  
DB 2221 PQVNNPKEMQLQVDFQKTMKVTVGTTQGVKSLTSMYKKEFLISSQDGHQWTLFFQNGKV 2280  
QY 1387 KVFQGNQDSFTPVNLSLDPDLLRYLRIHQSWHQTALRMEVLGCEAODLY 1438  
DB 2281 KVFQGNQDSFTPVNLSLDPDLLRYLRIHQSWHQTALRMEVLGCEAODLY 2332

RESULT 9

US-08-474-503-2  
; Sequence 2, Application US/08474503  
; Patent No. 5744446  
; GENERAL INFORMATION:  
; APPLICANT: Emory University  
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kilpatrick & Cody  
; STREET: 1100 Peachtree Street, Suite 2800  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: US  
; ZIP: 30309  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/474,503  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pratt, John S.  
; REGISTRATION NUMBER: 29,476  
; REFERENCE/DOCKET NUMBER: EMU106CIP(3)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 404-815-6500  
; TELEFAX: 404-815-6555  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2332 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapien  
; TISSUE TYPE: Liver cdna sequence  
US-08-474-503-2

Query Match 94.1%; Score 7234; DB 1; Length 2332;  
Best Local Similarity 61.7%; Pred. No. 0;  
Matches 1438; Conservative 0; Mismatches 0; Indels 894; Gaps 1;  
QY 1 ATRRYILGAVELSDWYMQSDIGELPVDARPPRPVKSPFNTSVVYKTLFVEFTVHLFN 60  
DB 1 ATRRYILGAVELSDWYMQSDIGELPVDARPPRPVKSPFNTSVVYKTLFVEFTVHLFN 60  
QY 61 IAKPRPPWMLGPTIOAEVYDVTITLKNMASHPVSLHVGVSYWKASGAEYDDQTSQ 120

DB 61 IAKPRPPWMLGPTIOAEVYDVTITLKNMASHPVSLHVGVSYWKASGAEYDDQTSQ 120  
QY 121 REKEDDKVPPGGSHTYVWQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 180  
DB 121 REKEDDKVPPGGSHTYVWQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 180  
QY 181 EGSILAKKCTQLHKPILLFAVDEGKSWHSETKNSLMODBDAAASARAWPKMHTVNGVYNR 240  
DB 181 EGSILAKKCTQLHKPILLFAVDEGKSWHSETKNSLMODBDAAASARAWPKMHTVNGVYNR 240  
QY 241 SLPGILGCHRKSVYWHVIGMTTPEVHSIFLEGHTEFLVRNHRQASLEISITTLTQTLTLL 300  
DB 241 SLPGILGCHRKSVYWHVIGMTTPEVHSIFLEGHTEFLVRNHRQASLEISITTLTQTLTLL 300  
QY 301 MDLGQFLLCFCHISSHQHDGMEAYVKVDSCEPEEPQLRMKNNEEAEDDDDLTDSMDVVR 360  
DB 301 MDLGQFLLCFCHISSHQHDGMEAYVKVDSCEPEEPQLRMKNNEEAEDDDDLTDSMDVVR 360  
QY 361 DDNSPSFTQIRSVAKKPKTWVHYIAAEEEDWDYAPLVLPDDRYSKQYLNNQFORIG 420  
DB 361 DDNSPSFTQIRSVAKKPKTWVHYIAAEEEDWDYAPLVLPDDRYSKQYLNNQFORIG 420  
QY 421 RYKVKRFMAYTDTETFKTREAIQHESGILGPLLYGVGDTLLIIFKNQASRPNIYPHGI 480  
DB 421 RYKVKRFMAYTDTETFKTREAIQHESGILGPLLYGVGDTLLIIFKNQASRPNIYPHGI 480  
QY 481 TDVRPLYSRRLPKGVKHLXDFPILPGEIFKYKWTVTVEGPTKSDPRCLTRYSSVFVME 540  
DB 481 TDVRPLYSRRLPKGVKHLXDFPILPGEIFKYKWTVTVEGPTKSDPRCLTRYSSVFVME 540  
QY 541 RDLASGLIGPLLI CYKESVDQGNQIMSKRNVLFSVFDENRSWLTENIQRLPNPAG 600  
DB 541 RDLASGLIGPLLI CYKESVDQGNQIMSKRNVLFSVFDENRSWLTENIQRLPNPAG 600  
QY 601 VQLEDPEFOASNTMHSINGVYVDSLOLSVCLHVAWYVYILSIGAQTDFLSVFTSGYTFKH 660  
DB 601 VQLEDPEFOASNTMHSINGVYVDSLOLSVCLHVAWYVYILSIGAQTDFLSVFTSGYTFKH 660  
QY 661 KMVYEDTLTLFPFSGETVFMSENPGIWIIGCHNSDFRNRGTMALLKVVSCDKNYGYE 720  
DB 661 KMVYEDTLTLFPFSGETVFMSENPGIWIIGCHNSDFRNRGTMALLKVVSCDKNYGYE 720  
QY 721 DSYEDISAYLLSKNNAIEPRSF 742  
DB 721 DSYEDISAYLLSKNNAIEPRSFQNGRHPSTROQFNATTIPENDIEKTDPFWAHRTPMP 780  
QY 743 742  
DB 781 KIQNVSSDLLMLLRSOPPHGLSLDLOEAKYETPDDPSGAIDSNNSLSWTHFRPQ 840  
QY 743 742  
DB 841 LHHSGDMVFTPESGLQLRLNEKLGTTAATLKKLDFKVSSTNNLSTIPSDNLAAGTDN 900  
QY 743 742  
DB 901 TSSLGPPSPVHYDSQLDITLFGKKSPLTESGGPLSENNDSKILSEGLMNSQBSW 960  
QY 743 742  
DB 961 GKNVSTESGRLPFGKRAHGAPALLTKDNALFKVSIISLLKTNKTSNNSATNRKTHIDGPSL 1020  
QY 743 742  
DB 1021 LIENSPVWQNIILSDTEPKVTPLIHDRMLDKNATALLNHNMSKNTSSKNMVMVQK 1080  
QY 743 742  
DB 1081 KEGPIPPDAQNPDMSPFKMLFLPESARWIQRTHGKNSLNSGGQSPSKQLVSLGPEKSV 1140  
QY 743 742  
DB 1141 QNFLSEKNVYVVGKEFTKDVGLKEMVFPSPSRNLFTNLNHNHNTNHNQEKIQBIEK 1200

QY 743 ----- 742  
Db 1201 KETLIQENVVLPOIHVVTGKNFKNLFLSTRQVSGYEGAYAPVLQDRSLNDSTNR 1260  
QY 743 ----- 742  
Db 1261 TKKHTAHFKKBEENLEGLNQTQKIVEKYACTTRISPNSTSQNFVORSKRALKQFRL 1320  
QY 743 ----- 742  
Db 1321 PLEETELEKRIIVDDTSTQSKNKKHLTPSTLTQIDYNEKEKGAITOSPLSDCLTRSHSI 1380  
QY 743 ----- 742  
Db 1381 PQANRSPPIAKVSSFFSIRPIYLTRVLFDQNSSHLPAASVRKXDSGVQSSHFLQAKK 1440  
QY 743 ----- 742  
Db 1441 NNLISLAILTLEMTGQREVGLGTATNSVTYKKVENTVLPKPLPKTSKGVKLLPKVHI 1500  
QY 743 ----- 742  
Db 1501 YKOLFPTETSGSPGHLDLVEGSLLOQTEGAIKWNEANRPGKVPFLVATESAKTPSK 1560  
QY 743 ----- 742  
Db 1561 LLDPLAWNHYGTQIPKEWKSQKSPKTAFFKKKDTILSLNACESHAIANAINEQONK 1620  
QY 743 -----  
Db 1621 EIEVTWAKQRTLRCSQNPVLRKHQREIRTRTLQSDQEEIDYDDTISVEMKXEDFDYI 1680  
QY 786  
Db 787 DEDENQSPRSQKTRHYFIAAVERLDYWGSSPHVLNRNQAQSGVQFKKVVQFETD 846  
QY 1740  
Db 847 GSFTQPLRGELNEHLGGLGPYIRAEVDNIWTFERNQASRPYSYSLISVEEDQROGA 906  
Db 1741 GSFTQPLRGELNEHLGGLGPYIRAEVDNIWTFERNQASRPYSYSLISVEEDQROGA 1800  
QY 907 EPRKNFVFNPTKTYFWKVQHMAPTKDFCKAWAYFSDVLEKDVHSGSLIGPLLVCHT 966  
Db 1801 EPRKNFVFNPTKTYFWKVQHMAPTKDFCKAWAYFSDVLEKDVHSGSLIGPLLVCHT 1860  
QY 967 NTLNPAHGRQVTVQBFALFTTIFDETKSWYFTENMERCEAPCNLOMEDPTEKENVRFA 1026  
Db 1861 NTLNPAHGRQVTVQBFALFTTIFDETKSWYFTENMERCEAPCNLOMEDPTEKENVRFA 1920  
QY 1027 INGYIMDTLPGLVMAQDQIRWYLLSMGNSNENIHSIFSGHVFTVRKKEEYKXALYNLYP 1086  
Db 1921 INGYIMDTLPGLVMAQDQIRWYLLSMGNSNENIHSIFSGHVFTVRKKEEYKXALYNLYP 1980  
QY 1087 GVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPGLMASGHIRDQITAS 1146  
Db 1981 GVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPGLMASGHIRDQITAS 2040  
QY 1147 GOYGQWAPKLARIHSGSINASTKEPFSWIKVDLLAPMIHGIKTQGARQKFSYISQ 1206  
Db 2041 GOYGQWAPKLARIHSGSINASTKEPFSWIKVDLLAPMIHGIKTQGARQKFSYISQ 2100  
QY 1207 FIIMYSLDGKCKTYRGNSTGTLMVFPNGVNDSSGKHNIFNPPIIARYIRLHPHTHSIRS 1266  
Db 2101 FIIMYSLDGKCKTYRGNSTGTLMVFPNGVNDSSGKHNIFNPPIIARYIRLHPHTHSIRS 2160  
QY 1267 TLRMELMGCDLNSCSPMLGMSKAISSAQITASSYFTNPFATWSPSKARLHLQGRNANR 1326  
Db 2161 TLRMELMGCDLNSCSPMLGMSKAISSAQITASSYFTNPFATWSPSKARLHLQGRNANR 2220  
QY 1327 PQVNNPKMQLQVDFQKTMKVTGTTQGVKSLLTSMYKPEFLISSODGHQWTLFFONGKV 1386  
Db 2221 PQVNNPKMQLQVDFQKTMKVTGTTQGVKSLLTSMYKPEFLISSODGHQWTLFFONGKV 2280

QY 1387 KYVQGNQDSFTPVNSLDPPLTRYLRIHQSWHQAIALRMEVLGCEAODLY 1438  
Db 2281 KYVQGNQDSFTPVNSLDPPLTRYLRIHQSWHQAIALRMEVLGCEAODLY 2332

RESULT 10  
US-08-670-707A-2  
; Sequence 2, Application US/08670707A  
; Patent No. 5859204  
; GENERAL INFORMATION:  
; APPLICANT: Lollar, John S.  
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII  
; NUMBER OF SEQUENCES: 40  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.  
; STREET: 5370 Manhattan Circle Suite 201  
; CITY: Boulder  
; STATE: Colorado  
; COUNTRY: USA  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA: US/08/670,707A  
; FILING DATE: 26-JUN-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/US94/13200  
; FILING DATE: 15-NOV-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/212,133  
; FILING DATE: 11-MAR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/864,004  
; FILING DATE: 07-APR-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Greenlee, Lorraine L.  
; REGISTRATION NUMBER: 27,894  
; REFERENCE/DOCKET NUMBER: 75-95F  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 303/499-8080  
; TELEFAX: 303/499-8089  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2332 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; TISSUE TYPE: Liver  
; US-08-670-707A-2

Query Match 94.1%; Score 7234; DB 2; Length 2332;  
Best Local Similarity 61.7%; Pred. No. 0;  
Matches 1438; Conservative 0; Mismatches 0; Indels 894; Gaps 1;

QY 1 ATRRYTLGAVELSWDMQSDLGELPVDARFPFRPKSPFNTSVVYKKTLPVEFTVHLFN 60  
Db 1 ATRRYTLGAVELSWDMQSDLGELPVDARFPFRPKSPFNTSVVYKKTLPVEFTVHLFN 60  
QY 61 IAKPRPPMGLLGPTIOAEVYDVTITLKNMASHVPSLHAGVSVYKASGAEYDDQTSQ 120  
Db 61 IAKPRPPMGLLGPTIOAEVYDVTITLKNMASHVPSLHAGVSVYKASGAEYDDQTSQ 120  
QY 121 REKEDDKVPFGSHYTVQVLKENGPMASDPLCLTYSYLSLHVDLVKDLNSGLIGALLVCR 180

Db	121	REXEDDKVFGGSHYYVQVWKGNGPMASDPLCLTITSYLSHVDDLVKDNLNGLGALLVCR	180
Qy	181	EGSLAKEKTQTLHKFILLFAVFDGKSWHSETKNSLMQDRDAASARAWPKOHTVNGYVNR	240
Db	181	EGSLAKEKTQTLHKFILLFAVFDGKSWHSETKNSLMQDRDAASARAWPKOHTVNGYVNR	240
Qy	241	SLPGLIGCHRSKSVYVHWVGMGTTPVEVHSIFLEGHTFLVRNHRQASLEISPIITFLTAQTLL	300
Db	241	SLPGLIGCHRSKSVYVHWVGMGTTPVEVHSIFLEGHTFLVRNHRQASLEISPIITFLTAQTLL	300
Qy	301	MDLGOFLLFCHISHOHDGMEAYVYKVDSCPEEPQLRMKNEEAEYDDDLTDSMDVVR	360
Db	301	MDLGOFLLFCHISHOHDGMEAYVYKVDSCPEEPQLRMKNEEAEYDDDLTDSMDVVR	360
Qy	361	DDNSPSFIQIRSVAKKHPTWVHYIAAEEDMDYAPLVAPDDRSYKSOYLNNGPQIRIG	420
Db	361	DDNSPSFIQIRSVAKKHPTWVHYIAAEEDMDYAPLVAPDDRSYKSOYLNNGPQIRIG	420
Qy	421	RKYKVRFMAYTDTFTKTREAIQHSIGILGPLLYGEVGDITLLIIFKNQASRPYNIYPHGI	480
Db	421	RKYKVRFMAYTDTFTKTREAIQHSIGILGPLLYGEVGDITLLIIFKNQASRPYNIYPHGI	480
Qy	481	TDVRLYSRRLPKGVKHLKDPPIIPGELFKYKWTVTVEDGPKSDPCLTRYSSFYVME	540
Db	481	TDVRLYSRRLPKGVKHLKDPPIIPGELFKYKWTVTVEDGPKSDPCLTRYSSFYVME	540
Qy	541	RDLASGLIGPLLIICYESVDQRGNQIMSDKENVILFSVFDENRSWYLTIENIQRLPNPAG	600
Db	541	RDLASGLIGPLLIICYESVDQRGNQIMSDKENVILFSVFDENRSWYLTIENIQRLPNPAG	600
Qy	601	VOLEDPFOASNMHSINGVYFDSLSQVCLHEVAYWYILSIGAQDTFLSVFFSGYTFKH	660
Db	601	VOLEDPFOASNMHSINGVYFDSLSQVCLHEVAYWYILSIGAQDTFLSVFFSGYTFKH	660
Qy	661	KWYEDTTLTPFSGEYVFMENPGWLILGCHNSDFRNGMTALLKVSSCDKNTGDIYE	720
Db	661	KWYEDTTLTPFSGEYVFMENPGWLILGCHNSDFRNGMTALLKVSSCDKNTGDIYE	720
Qy	721	DSYEDISAYLLSKNNAIEPRSF-	742
Db	721	DSYEDISAYLLSKNNAIEPRSFQNSRHPSTRQKQFNATTIPENDIEKTPFAHRTMP	780
Qy	743	-----	742
Db	781	KIQNVSSDLLMLLRQSPTPHGLSLSDIQEAKYETFFSDPSPGALDSNNSLSEMTHERPQ	840
Qy	743	-----	742
Db	841	LHSGDMVFTPESGIQLRLNEKLGTTAATELKGCLDFKVSSTSNLITSPSDNLAAGTDN	900
Qy	743	-----	742
Db	901	TSSLGPPSPMVHYDSQDITTLFGKSSPLTESGGPLSEENNDKLLSGLMNSQBSW	960
Qy	743	-----	742
Db	961	GKNVSTESGRLFKGKRAHPALLTKDNALFKVSI SLLTKNTKTSNNSATNRKTHDGPSL	1020
Qy	743	-----	742
Db	1021	LIENSPSVQNI LESTDBFKKVTPLIHDRMLMDKNATRLNHNMSNKTTSKKNMWOQK	1080
Qy	743	-----	742
Db	1081	KEGP1PPDAQNPDMSFFKMLFLPESARWIQRTGKNSLNSGQSPSPKQVLSLGPBKSVEG	1140
Qy	743	-----	742
Db	1141	QNFLSEKNVVGKGEFTKDVGLKEMVFPSSRNFLTNLDNLHNHNTNQBKQIEBIEK	1200
Qy	743	-----	742

1201	KETLIQENVVLPOIHTVTGTGNPKMKNLFLLSLSTRQNVESGYEGAYAPVLQDFRSLNDSTNR	1260
743	-----	742
1261	TKKHTAHFSGKGBEENLEGLGNQTKOIVEKYACTTRISPNTSQONFVTOQRSEKALQFRL	1320
743	-----	742
1321	PLEBTELEKRIIVDDTSTQWSKMKHLPSTLTQIDYNEKEKGAITQSPUSDCLTRSHSI	1380
743	-----	742
1381	PQANRSPLPKAVSSPPSIRPIYLTRVLFDQDNSSHLPAASRYKKDKSGVBSHFLQGAKX	1440
743	-----	742
1441	NNLSLAILTLEMTGDQREVGSLGTSATNSVYKKVENTVLPKPDLPKTSKGVKELLPKVHI	1500
743	-----	742
1501	YQKDLPTTSTNGSPGHLDLVEGSLLOGTBGAIKWNEANRPKVPFLRVATESAKTPSK	1560
743	-----	742
1561	LLDPLAWNHYGTQIPKEEWSQESKPEKTAFFKKDDTILSNACSNHATAINEGQNK	1620
743	-----SONPPVLKXHQRETRTRTLQSDQBEIDYDDDTISYEMKKEDDIY	786
1621	EIEVTWAKQRTERLCSNPPVLKXHQRETRTRTLQSDQBEIDYDDDTISYEMKKEDDIY	1680
787	DEBNGSPRSFOKTRHYFTAAVERLWDYQWSSSPHVLNRNRAQSGSVQPKVVVFQFTD	846
1681	DEBNGSPRSFOKTRHYFTAAVERLWDYQWSSSPHVLNRNRAQSGSVQPKVVVFQFTD	1740
847	GSFTQPLRYGELNEHLGLGPIYIRAEVEDNIMVTFRNQASRPYSFYSSLSIYEEDQOGA	906
1741	GSFTQPLRYGELNEHLGLGPIYIRAEVEDNIMVTFRNQASRPYSFYSSLSIYEEDQOGA	1800
907	EPRKNFVKNETKTYFWKVOHMAPTKDEPDCAKAWAYFSDVDLEKOVHSGLIGLPLVCHT	966
1801	EPRKNFVKNETKTYFWKVOHMAPTKDEPDCAKAWAYFSDVDLEKOVHSGLIGLPLVCHT	1860
967	NTLNPAGROVTVQEFALFTTIDETKSWYFTENMERNCRAPCNIOMEDPTFKENYFHA	1026
1861	NTLNPAGROVTVQEFALFTTIDETKSWYFTENMERNCRAPCNIOMEDPTFKENYFHA	1920
1027	INGYIMDTLPGLVMAQOQIRWYLLSMGNSNENIHSIHFSGHVFTVRKKEEYKMALYNLYP	1086
1921	INGYIMDTLPGLVMAQOQIRWYLLSMGNSNENIHSIHFSGHVFTVRKKEEYKMALYNLYP	1980
1087	GVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPLGMASGHIRDFOQTAS	1146
1981	GVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPLGMASGHIRDFOQTAS	2040
1147	GOYQWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMIHGIKTQAGOKFSSLYISQ	1206
2041	GOYQWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMIHGIKTQAGOKFSSLYISQ	2100
1207	FIIMYSLDGKKQWYRGNSGTGLMVFFGNVDSGIGKINIFNPPIIARYIRLPHYSIRS	1266
2101	FIIMYSLDGKKQWYRGNSGTGLMVFFGNVDSGIGKINIFNPPIIARYIRLPHYSIRS	2160
1267	TLRMLMGCDLNSCMPGLMESKASIDAQITASSYFTNMFTATWSPSKARLHLOGRSNWR	1326
2161	TLRMLMGCDLNSCMPGLMESKASIDAQITASSYFTNMFTATWSPSKARLHLOGRSNWR	2220
1327	PQVNNPKWLOVDPKTKMKTGVTTQGVKSLTSMYKBEFLISSQGHQWTLFPQNGKV	1386
2221	PQVNNPKWLOVDPKTKMKTGVTTQGVKSLTSMYKBEFLISSQGHQWTLFPQNGKV	2280
1387	KVFGQNDSPFPVNVSLDPLLTLYRLIHPOSWHQIALRMEVLGCEAODLY	1438
2281	KVFGQNDSPFPVNVSLDPLLTLYRLIHPOSWHQIALRMEVLGCEAODLY	2332

Wed Dec 10 11:56:46 2003

RESULT 11

US-09-037-601-2  
; Sequence 2, Application US/09037601  
; Patent No. 6180371  
; GENERAL INFORMATION:  
; APPLICANT: Lollar, John S.  
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII  
; NUMBER OF SEQUENCES: 40  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.  
; STREET: 5370 Manhattan Circle Suite 201  
; CITY: Boulder  
; STATE: Colorado  
; COUNTRY: USA  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/037,601  
; FILING DATE: 26-JUN-1996  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/US94/13200  
; FILING DATE: 15-NOV-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/212,133  
; FILING DATE: 11-MAR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/864,004  
; FILING DATE: 07-APR-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ferber, Donna M.  
; REGISTRATION NUMBER: 33,878  
; REFERENCE/DOCKET NUMBER: 75-95F  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 303/499-8089  
; TELEFAX: 303/499-8089  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2332 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: protein  
; HYPOTHEICAL: YES  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; TISSUE TYPE: Liver  
; US-09-037-601-2

Query Match 94.1%; Score 7234; DB 3; Length 2332;  
Best Local Similarity 61.7%; Pred. No. 0;  
Matches 1438; Conservative 0; Mismatches 0; Indels 894; Gaps 1;

Qy 1 ATRRYIYGAVELSDWYMSDGLGELPVDARPPRPVKSPFNTSVVYKTLFVEFTVHLFN 60  
Db 1 ATRRYIYGAVELSDWYMSDGLGELPVDARPPRPVKSPFNTSVVYKTLFVEFTVHLFN 60  
Qy 61 IAKRPPWMLLGGTIIQAEVYDVTVITLKNASHPVSLHVGVSYWKASGAEYDDQTSQ 120  
Db 61 IAKRPPWMLLGGTIIQAEVYDVTVITLKNASHPVSLHVGVSYWKASGAEYDDQTSQ 120  
Qy 121 REKEDDKVPGSGSHYVQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 180  
Db 121 REKEDDKVPGSGSHYVQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 180

Qy 181 EGSIAKEKTOIHKFILLFAVDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGVYNR 240  
Db 181 EGSIAKEKTOIHKFILLFAVDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGVYNR 240  
Qy 241 SLPLIGCHRSVYVHVIGMGTTPVEHISIFLEGHTFLVRNHRQASLSIPITFLTAQTL 300  
Db 241 SLPLIGCHRSVYVHVIGMGTTPVEHISIFLEGHTFLVRNHRQASLSIPITFLTAQTL 300  
Qy 301 MDLQOFLFCHISSHQHDGMEAYVYKVDSCPEEPQLRMKNNEAEADYDDDLTDSMDVVRP 360  
Db 301 MDLQOFLFCHISSHQHDGMEAYVYKVDSCPEEPQLRMKNNEAEADYDDDLTDSMDVVRP 360  
Qy 361 DDNSPSFIQIRSVAKKHPKTHVHIAAEEDWDYAPLVLPADDRSKYSQYLNNGPQIRG 420  
Db 361 DDNSPSFIQIRSVAKKHPKTHVHIAAEEDWDYAPLVLPADDRSKYSQYLNNGPQIRG 420  
Qy 421 RYKVKRFMAYTDTFTKTRAIQHESGILGLPGLLYGEVGDITLLIIFKQASRPYNIYPHGI 480  
Db 421 RYKVKRFMAYTDTFTKTRAIQHESGILGLPGLLYGEVGDITLLIIFKQASRPYNIYPHGI 480  
Qy 481 TDVRLYSRRLPKGVKHLKOPPIIPGELFKYKWTVTVEDGPKSDPRCLTRYYSFVWME 540  
Db 481 TDVRLYSRRLPKGVKHLKOPPIIPGELFKYKWTVTVEDGPKSDPRCLTRYYSFVWME 540  
Qy 541 RDLASGLIGPLLI CYKESVDQGRNQIMSDKENVILFSVFDNRNRSWYLTENQRFILNPAG 600  
Db 541 RDLASGLIGPLLI CYKESVDQGRNQIMSDKENVILFSVFDNRNRSWYLTENQRFILNPAG 600  
Qy 601 VQLEDPEFQASNIHMSINGVYVDFSIQLSVCLHEVAYWYILSIGAQDTFLSFFSGYTPKH 660  
Db 601 VQLEDPEFQASNIHMSINGVYVDFSIQLSVCLHEVAYWYILSIGAQDTFLSFFSGYTPKH 660  
Qy 661 KMVEDTLTLFPFSGETVFMENPGLMILCHNSDFRNQMTALLKVVSCDKNTGDYVE 720  
Db 661 KMVEDTLTLFPFSGETVFMENPGLMILCHNSDFRNQMTALLKVVSCDKNTGDYVE 720  
Qy 721 DSYEDISAYLISKNNAIERPSF----- 742  
Db 721 DSYEDISAYLISKNNAIERPSFQNSRHPSTRQKQFNATTPENDIEKTDPFPAHRTMP 780  
Qy 743 ----- 742  
Db 781 KIQNVSSDLLMLLRQSPTPHGLSLDLQAEKYETFSDDPSPGAIDSNNSLSEMTHERPQ 840  
Qy 743 ----- 742  
Db 841 LHSQDMVFTPEGLQLRLNEKLGTTAATELKKLDFKVSSTSNLIITIPSDNLAAGTDN 900  
Qy 743 ----- 742  
Db 901 TSSLGPPSPVHYDSQDITTLFGKXSSPLTESGGPLSLEENNDKLSGLMSQESSW 960  
Qy 743 ----- 742  
Db 961 GKNVSTESGRLEFKGKRAHPALLTKDNALFKVISLTKNTKNSNATNRKTHIDGPSL 1020  
Qy 743 ----- 742  
Db 1021 LIENSPVQNILESDETFKVTPLIHDRMLADKNATALRLNHSNKTTSKNNMEMVQOK 1080  
Qy 743 ----- 742  
Db 1081 KEGPIPPDAQNPDMSPFKMLFLPESARWIORTHGKNSLNSGGQSPKQLVSLGPEKSVEG 1140  
Qy 743 ----- 742  
Db 1141 QNFLSEKNVVKVGKGFYKDVGLKEMVFPSSRNLFLTNDLNLHNTHNQEKIKQIEIEK 1200  
Qy 743 ----- 742  
Db 1201 KETLIQENVVLPOIHTVTGTGKNFMKNLFLLSSTRQNVESYEGAYAPVLQDFRSLNDSTNR 1260  
Qy 743 ----- 742



Db 1261 TKKHTAHFSKKGEENLEBGLNQTQKQIVKIACTTRISNTSOQNFTQORSKALKQFRL 1320  
QY 743 ----- 742  
Db 1321 PLEETELEKRIIVDDTSTQSKNMKHLTPSTLTQIDYNEKEGAIQSPLSDCLTRSHSI 1380  
QY 743 ----- 742  
Db 1381 PQANRSPLEIAKVSSPFSIRPIYLTVLFQDNSSHLPAASRYRKDSGVQESSHFLQAKK 1440  
QY 743 ----- 742  
Db 1441 NNLSLAILTEMTGDQREVGSLGTSATNSVYKKVENTVLPKPDLPKTSKVELLPKVHI 1500  
QY 743 ----- 742  
Db 1501 YQXDLPTETSNQSPGHLDLVEGSLLOQTEGAIKWNEANRPKVPFLRVATESSAKTPSK 1560  
QY 743 ----- 742  
Db 1561 LLDPLAWNHYQTQIPKEEWSQESPEKTAFFKKKDTILSLNACESNHAIAINEQONKP 1620  
QY 743 ----- 786  
Db 1621 EIEVTWAKQGRBELCSQNPVLKRHQREITRTTLOSQDEIDYDDTISVEMKKEDFDIY 1680  
QY 787 DEBENOSPRGFKTRHYFLAAVERLDYQWSSPHVLRNRAQSGSVQPKKVVFOBFDT 846  
Db 1681 DEBENOSPRGFKTRHYFLAAVERLDYQWSSPHVLRNRAQSGSVQPKKVVFOBFDT 1740  
QY 847 GSTQPLRGELNEHLLGPGYIRAEVEDNIMVTFRNQASRPYSFYSSLSIYSEEDQOQA 906  
Db 1741 GSTQPLRGELNEHLLGPGYIRAEVEDNIMVTFRNQASRPYSFYSSLSIYSEEDQOQA 1800  
QY 907 EPRKNFVKNETKTYFWKVOHMAPKYDEDFCKAWAYFSDVLEKDVHSGILGPLLVCYT 966  
Db 1801 EPRKNFVKNETKTYFWKVOHMAPKYDEDFCKAWAYFSDVLEKDVHSGILGPLLVCYT 1860  
QY 967 NTLNPAHGRQVTVQEBALFTTIDETKSWYFTENMERNCRAPCNIOMEDPTFKENTRFHA 1026  
Db 1861 NTLNPAHGRQVTVQEBALFTTIDETKSWYFTENMERNCRAPCNIOMEDPTFKENTRFHA 1920  
QY 1027 INGYIMDTPLGLVMAQDQIRWYLLSGNSNENIHSIHFSGHVFTVRKKEEYKMAYNLYP 1086  
Db 1921 INGYIMDTPLGLVMAQDQIRWYLLSGNSNENIHSIHFSGHVFTVRKKEEYKMAYNLYP 1980  
QY 1087 GVPETVEMLPKAGIWRVECLIGEHLHAGMSTFLVYSNKCQTPPLGNASGHIRDFOITAS 1146  
Db 1981 GVPETVEMLPKAGIWRVECLIGEHLHAGMSTFLVYSNKCQTPPLGNASGHIRDFOITAS 2040  
QY 1147 GOYGOWAPKLARLHYSGSINAWSTKBPFSWKVDLLAPMIIGHIKTGARQKFSLLYISQ 1206  
Db 2041 GOYGOWAPKLARLHYSGSINAWSTKBPFSWKVDLLAPMIIGHIKTGARQKFSLLYISQ 2100  
QY 1207 FIIMYSIDGKKWQTYRGNSGTGLMVFFGNVDSSGIKENIENPPIIARYIRLHPHYSIRS 1266  
Db 2101 FIIMYSIDGKKWQTYRGNSGTGLMVFFGNVDSSGIKENIENPPIIARYIRLHPHYSIRS 2160  
QY 1267 TLRMELMGCDLNSCSMPGLMESKAISDAQITASSYFTNMPATWSPSKARLHLQGRSNWR 1326  
Db 2161 TLRMELMGCDLNSCSMPGLMESKAISDAQITASSYFTNMPATWSPSKARLHLQGRSNWR 2220  
QY 1327 POWNPKEMLOVDPOKTMKYTGVTQGVKSLITSMYKFEFLISSQDGHQWTLFPONGKV 1386  
Db 2221 POWNPKEMLOVDPOKTMKYTGVTQGVKSLITSMYKFEFLISSQDGHQWTLFPONGKV 2280  
QY 1387 KVFQGNQDSFTPVVNSLDPPLLTRYLRIHPQSWHQAIALRMEVLGCEAQDLY 1438  
Db 2281 KVFQGNQDSFTPVVNSLDPPLLTRYLRIHPQSWHQAIALRMEVLGCEAQDLY 2332

US-09-315-179-2  
; Sequence 2, Application US/09315179  
; Patent No. 6376463  
; GENERAL INFORMATION:  
; APPLICANT: Lollar, John S  
; TITLE OF INVENTION: Modified Factor VIII  
; FILE REFERENCE: 75-95H  
; CURRENT APPLICATION NUMBER: US/09/315,179  
; CURRENT FILING DATE: 1999-05-20  
; EARLIER APPLICATION NUMBER: U.S. 09/037,601  
; EARLIER FILING DATE: 1998-03-10  
; EARLIER APPLICATION NUMBER: U.S. 08/670,707  
; EARLIER FILING DATE: 1996-06-26  
; EARLIER APPLICATION NUMBER: PCT/US97/111155  
; EARLIER FILING DATE: 1997-06-26  
; EARLIER APPLICATION NUMBER: PCT/US94/13200  
; EARLIER FILING DATE: 1994-11-15  
; EARLIER APPLICATION NUMBER: U.S. 08/212,133  
; EARLIER FILING DATE: 1994-03-11  
; EARLIER APPLICATION NUMBER: U.S. 07/864,004  
; EARLIER FILING DATE: 1992-04-07  
; NUMBER OF SEQ ID NOS: 40  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 2332  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-315-179-2

Query Match 94.1%; Score 7234; DB 4; Length 2332;  
Best Local Similarity 61.7%; Pred. No. 0;  
Matches 1438; Conservative 0; Mismatches 0; Indels 894; Gaps 1;

QY 1 ATRYYVLGAVELSWDMQSDLGELPVDARFPVRPKSPFNTSVVYKTLFVEFTVHLFN 60  
Db 1 ATRYYVLGAVELSWDMQSDLGELPVDARFPVRPKSPFNTSVVYKTLFVEFTVHLFN 60  
QY 61 IAKPRPPNMLGPTTQAEVYDVTVTITKNMASHPVSLHAVGVSYKASGABYDDQTSQ 120  
Db 61 IAKPRPPNMLGPTTQAEVYDVTVTITKNMASHPVSLHAVGVSYKASGABYDDQTSQ 120  
QY 121 REKEDDKVPFGSSHTYVQVLKENGPMASDPLCLTYSLSHVDLVKDLNSGLIGALLVCR 180  
Db 121 REKEDDKVPFGSSHTYVQVLKENGPMASDPLCLTYSLSHVDLVKDLNSGLIGALLVCR 180  
QY 181 EGSIAKEKTOTLHKFTLLFAVDFDEGKSWESETKNSLMQDDAASARAWPQHTVGVYNR 240  
Db 181 EGSIAKEKTOTLHKFTLLFAVDFDEGKSWESETKNSLMQDDAASARAWPQHTVGVYNR 240  
QY 241 SLPLGIGCHRSYVYHVIWGTTPPEVHSIFLEGHTFLVRNHRQASLEISPTITLTAQTLL 300  
Db 241 SLPLGIGCHRSYVYHVIWGTTPPEVHSIFLEGHTFLVRNHRQASLEISPTITLTAQTLL 300  
QY 301 MDLQGFLLCHTSSHQHDCMEAYKVYDSCPEEPQLEMKNEEAEDYDDLTDSEMDVVRP 360  
Db 301 MDLQGFLLCHTSSHQHDCMEAYKVYDSCPEEPQLEMKNEEAEDYDDLTDSEMDVVRP 360  
QY 361 DDNSPSFTQIRSVAKKHPTWVHYIAAEEEDMDYAPVLAPDDRYSKSYOLNNGPQIRG 420  
Db 361 DDNSPSFTQIRSVAKKHPTWVHYIAAEEEDMDYAPVLAPDDRYSKSYOLNNGPQIRG 420  
QY 421 RYKVKYRFMAYTDETFKTRERAIQHESGILGPLYGEVGDTLIIIPKQASRPYNIYPHGI 480  
Db 421 RYKVKYRFMAYTDETFKTRERAIQHESGILGPLYGEVGDTLIIIPKQASRPYNIYPHGI 480  
QY 481 TDVRPLYSRRLPKGVKHLKDPFLLPQRIPKYKWTVTVEGPTKSDPRCLTRYYSFVWME 540  
Db 481 TDVRPLYSRRLPKGVKHLKDPFLLPQRIPKYKWTVTVEGPTKSDPRCLTRYYSFVWME 540  
QY 541 RDLASGLIGPLLIICYKESVDORGNQJMSDKRNVILSFVDENRSWYLTENIOEFLPNPAG 600  
Db 541 RDLASGLIGPLLIICYKESVDORGNQJMSDKRNVILSFVDENRSWYLTENIOEFLPNPAG 600

QY 601 VQEDPEFQASNMHSINGVFDLSQVCLHEVAYWYILSIGAQDTFLSVFSGYTFKH 660  
Db 601 VQEDPEFQASNMHSINGVFDLSQVCLHEVAYWYILSIGAQDTFLSVFSGYTFKH 660  
QY 661 KMVEDTLTLPFSGETVFMSEMPGLWILCHNSDFRNRGWTALLKVSSCDKNTGDYIE 720  
Db 661 KMVEDTLTLPFSGETVFMSEMPGLWILCHNSDFRNRGWTALLKVSSCDKNTGDYIE 720  
QY 721 DSYEDISAYLLSKNAIEPRSF----- 742  
Db 721 DSYEDISAYLLSKNAIEPRSFQNSRHPSTRQOFNATIPENDIEKTDWFAHRTMP 780  
QY 743 ----- 742  
Db 781 KIQNVSSDLLMLLRQSPHPHGLSLDIAEAKYETFSDDPSGAIDSNLSLSEMTFRPQ 840  
QY 743 ----- 742  
Db 841 LHSQDMVFTPESGIQLRNEKLGTAATELKLDFKVSSTNNLISTIPSDNLAAGTDN 900  
QY 743 ----- 742  
Db 901 TSSLGPPSPMHHYDSQDITTLFGKSSPLTBSGGPLSLSEENDSKLLSEGLMNSQESSM 960  
QY 743 ----- 742  
Db 961 GKNVSTESGRLFKGRAHPALLTKONALFKVISLLKTKNTSNNATNRKTHIDGPSL 1020  
QY 743 ----- 742  
Db 1021 LIENSPVWONILESDBEFKVTPLIHRMLMDKNATALRLNHSNKTTSSKNMENVQOK 1080  
QY 743 ----- 742  
Db 1081 KEGPIPPDAQNPDMSFFKMLFLPESARWIORTHGKNSLNGOGSPKQLVSLGPEKSVBG 1140  
QY 743 ----- 742  
Db 1141 QNFLSEKNKVVGKBEFTKDVGLKEMVPPSRNLFNTLNDLHNHNTNHOEKKIOEIEK 1200  
QY 743 ----- 742  
Db 1201 KETLIQENVLPQIHTVTGKFMKMLFLLSTRQNVESYEGAYAPVLQDFRSLNDSTNR 1260  
QY 743 ----- 742  
Db 1261 TKKHTAHFSKKEENLEGLGNQTKOIVEKYACTTISPTSQNFVQTSKRALQKPL 1320  
QY 743 ----- 742  
Db 1321 PLEETELEKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKEGAIQSPSLDCLTRSHSI 1380  
QY 743 ----- 742  
Db 1381 PQANRSPFIKAVSSPFSIRPIYLTVLFDQNSHLPAAASYRKDSQVQESSHFLQAKK 1440  
QY 743 ----- 742  
Db 1441 NNLSLAILLEMTQDQREVGLSATSNTSVYKKVENTVLKPDLPKTSKVELLPKVHI 1500  
QY 743 ----- 742  
Db 1501 YQXDLFPETNSGPHGLDLVEGILLQTEGAIKWNEANRPGKVPFLRVATESAKTPSK 1560  
QY 743 ----- 742  
Db 1561 LLDPLAMDNHYGTQIPKEWKSQESPEKTAFAKKDTILSLNACESNHAIAANEQNK 1620  
QY 743 -----SONPVLKRHOREITRITLQSDQEEIDYDDTISVEMKEDPDYI 786  
Db 1621 EIEVTWAKQRTERLCSQNPVLRHOREITRITLQSDQEEIDYDDTISVEMKEDPDYI 1680  
QY 787 DEDENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNRNAQSGSVFPQKVVQBFETD 846

Db 1681 DEDENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNRNAQSGSVFPQKVVQBFETD 1740  
QY 847 GSTFTQPLXGELNEHLLGPGYIRAEVEDONIMVTFRNOASRPYSFSSLSIYSEDQROGA 906  
Db 1741 GSTFTQPLXGELNEHLLGPGYIRAEVEDONIMVTFRNOASRPYSFSSLSIYSEDQROGA 1800  
QY 907 EPRKNFVKNETTYFWKQOHMAPTKDEPDKAMAYFSDVLEKDVHSLIGLPLVCHT 966  
Db 1801 EPRKNFVKNETTYFWKQOHMAPTKDEPDKAMAYFSDVLEKDVHSLIGLPLVCHT 1860  
QY 967 NTLNPAHGRQVTVQEFALFTTIDETKSWYFTENMERNCRAPCNIOMEDTTFKENTYRPHA 1026  
Db 1861 NTLNPAHGRQVTVQEFALFTTIDETKSWYFTENMERNCRAPCNIOMEDTTFKENTYRPHA 1920  
QY 1027 INGYIMDTLPGLVMAQDQIRWYLLSMGNSNENIHSIFSGHVFTVRKKEBKALYNLYP 1086  
Db 1921 INGYIMDTLPGLVMAQDQIRWYLLSMGNSNENIHSIFSGHVFTVRKKEBKALYNLYP 1980  
QY 1087 GVPETVEMLPKAGIWEVECLIGBHLHAGMSTLPLVYSNKCOTPLGMASGHIRDFQITAS 1146  
Db 1981 GVPETVEMLPKAGIWEVECLIGBHLHAGMSTLPLVYSNKCOTPLGMASGHIRDFQITAS 2040  
QY 1147 GOYQWAPKLARLHYSINAWSTKBPFSWIKVDLLAPMIHGIKTQCARQKFSLSYISQ 1206  
Db 2041 GOYQWAPKLARLHYSINAWSTKBPFSWIKVDLLAPMIHGIKTQCARQKFSLSYISQ 2100  
QY 1207 FIIMYSLDGKKWQYRGNSTGTLNVFNGVNDSSGIGKNI FNPPIIARIYIRLHPHYSIRS 1266  
Db 2101 FIIMYSLDGKKWQYRGNSTGTLNVFNGVNDSSGIGKNI FNPPIIARIYIRLHPHYSIRS 2160  
QY 1267 TLRMELMGCDLNSCMLPGMESKALSDAQITASSYFTNMATWSPSKARLHLOQESNAR 1326  
Db 2161 TLRMELMGCDLNSCMLPGMESKALSDAQITASSYFTNMATWSPSKARLHLOQESNAR 2220  
QY 1327 PQVNNPKEWLQVDFQKTMKVTVGTQGVKSLTSMYKFEFLISSQDGHQWTLFFQNGKV 1386  
Db 2221 PQVNNPKEWLQVDFQKTMKVTVGTQGVKSLTSMYKFEFLISSQDGHQWTLFFQNGKV 2280  
QY 1387 KVFQNGQDSFTPVVNSLDPPLLTRYLRIHPOSWHQIALRMEVLGCEAODLY 1438  
Db 2281 KVFQNGQDSFTPVVNSLDPPLLTRYLRIHPOSWHQIALRMEVLGCEAODLY 2332

RESULT 13  
US-09-523-656-2  
; Sequence 2, Application US/09523656  
; Patent No. 6458563  
; GENERAL INFORMATION:  
; APPLICANT: Lollar S., John  
; TITLE OF INVENTION: MODIFIED FACTOR VIII  
; FILE REFERENCE: 75-951  
; CURRENT APPLICATION NUMBER: US/09/523,656  
; CURRENT FILING DATE: 2000-03-10  
; EARLIER APPLICATION NUMBER: 09/037,601  
; EARLIER FILING DATE: 1998-03-10  
; EARLIER APPLICATION NUMBER: 08/670,707  
; EARLIER FILING DATE: 1996-06-26  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 2332  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-523-656-2

Query Match 94.1%; Score 7234; DB 4; Length 2332;  
Best Local Similarity 61.7%; Pred. No. 0;  
Matches 1438; Conservative 0; Mismatches 0; Indels 894; Gaps 1;  
QY 1 ATRRYLGAVELSWDYMQSDLGELPVDARPPRPVKSPFNTSVYKTLFVEFTVHLFN 60  
Db 1 ATRRYLGAVELSWDYMQSDLGELPVDARPPRPVKSPFNTSVYKTLFVEFTVHLFN 60



```

Db      2221  PQVNNKEWLQVDFQKTKMVTGVTTCQKSLLSMVKBEFLISSQDGHQWTLFFQNGKV 2280
      1387  KVFQGNQDSFTPVVNSLDPPLLRYLRIHPQSWWHQIALRMEVLGCEAODLY 1438
      2281  KVFQGNQDSFTPVVNSLDPPLLRYLRIHPQSWWHQIALRMEVLGCEAODLY 2332

RESULT 14
PCT-US93-03275-4
; Sequence 4, Application PC/TUS9303275
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; APPLICANT: Runge, Marshall S.
; TITLE OF INVENTION: Hybrid Human/Porcine Factor VIII
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 1100 Peachtree Street, Suite 2800
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30309-4530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/03275
; FILING DATE: 19930407
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/864004
; FILING DATE: 07-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Rabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: EMU 106PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-815-6508
; TELEFAX: 404-815-6555
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; TISSUE TYPE: Liver cDNA sequence
PCT-US93-03275-4

Query Match . 94.1%; Score 7234; DB 5; Length 2332;
Best Local Similarity 61.7%; Pred.No. 0;
Matches 1438; Conservative 0; Mismatches 0; Indels 894; Gaps 1

QY      1  ATRRYYLGAVELSWDMQSDLGELPVDARPPRPVKSPFNTSVYVKTLFVEFTVHLFN 60
Db      1  ATRRYYLGAVELSWDMQSDLGELPVDARPPRPVKSPFNTSVYVKTLFVEFTVHLFN 60
QY      61  IAKPRPPWMLLGPTTCAEYVDVTWTLKNWASHPVLSHAGVSYWKASGEAYDDQTSQ 120
Db      61  IAKPRPPWMLLGPTTCAEYVDVTWTLKNWASHPVLSHAGVSYWKASGEAYDDQTSQ 120
QY      121  REKEDDKVFPGGSHYTVQVLEKNGPMASDPLCLITYSLSHVDLVKDLNSGLIGALLVCR 180
Db      121  REKEDDKVFPGGSHYTVQVLEKNGPMASDPLCLITYSLSHVDLVKDLNSGLIGALLVCR 180

```

```

Db      1261  TKKHTAHFKGGEENLEGLGNQTKQIVKEYACTTRISPTNSQCNFVQSRKALQKFL 1320
Qy      743  -----
Db      1321  PLEETELEKRIIVDDTSTQWKNMKHLTPSTLTQIDYNEKEGAIITQSPSLDCLTRSHSI 1380
Qy      743  -----
Db      1381  POANRSPPLIAKVSSPFSIRPIYLTVLFDQNSSHLPAASRYRKDGSGVQESSHFLQAKK 1440
Qy      743  -----
Db      1441  NNLSLAILFEMTGDOREVGSLGTSATNSVYKKVENTVLPRDLPKTSKGVKELLPKVHI 1500
Qy      743  -----
Db      1501  YQKDLFPFTSNGSPGHLDLVEGSLIQGTGGAIKWNEANRPKVPFLRVATESSAKTPSK 1560
Qy      743  -----
Db      1561  LLDPLAWNHNHGTQIPKEBWKQSKSPEKTAFFKKDTILSLNACESNHAIAINEGQNK 1620
Qy      743  -----
Db      1621  EIEVTVAKQGRTERLCSQNPVVLKRHQREITRTTLOSDDQEEIDYDDTISVEMKKEFDIY 1680
Qy      787  DEENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHLNRNRAQSGSVPOPKKVVQFQFTD 846
Db      1681  DEENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHLNRNRAQSGSVPOPKKVVQFQFTD 1740
Qy      847  GSFTQPLRGELNEHLGLGPIYRAVEDNIMVTFRNQASRPYSFYSSLSISYBEDQOQA 906
Db      1741  GSFTQPLRGELNEHLGLGPIYRAVEDNIMVTFRNQASRPYSFYSSLSISYBEDQOQA 1800
Qy      907  EPRKNFVKNETKTYFWKVOHMAPTKDPDCAMAYFSDVDLEKDVHSGSLIGPLVCHT 966
Db      1801  EPRKNFVKNETKTYFWKVOHMAPTKDPDCAMAYFSDVDLEKDVHSGSLIGPLVCHT 1860
Qy      967  NTLNPAHGRQVTVQEFALPFTTIDETKSWYFTENMERNCAPCNIOEMDPTFKENYRFA 1026
Db      1861  NTLNPAHGRQVTVQEFALPFTTIDETKSWYFTENMERNCAPCNIOEMDPTFKENYRFA 1920
Qy      1027  INGYIMDTPLGLVMAQDQRIWYLLSGNSNENIHSIHFGSHVFTVRKKEEYKMAIYNLYP 1086
Db      1921  INGYIMDTPLGLVMAQDQRIWYLLSGNSNENIHSIHFGSHVFTVRKKEEYKMAIYNLYP 1980
Qy      1087  GVFTVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPPLGWSGHIEDFOITAS 1146
Db      1981  GVFTVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPPLGWSGHIEDFOITAS 2040
Qy      1147  GOYGOWAPKLARLHYSGSINAWSTKBPFSWIKVDLLAPMIIHGIKTGQARQKFSYISQ 1206
Db      2041  GOYGOWAPKLARLHYSGSINAWSTKBPFSWIKVDLLAPMIIHGIKTGQARQKFSYISQ 2100
Qy      1207  FIIMYSLDGKKWQTVRGNSTGTLVWFFGNVDSGKIKHINFPPIIARIYIRLPHYSIRS 1266
Db      2101  FIIMYSLDGKKWQTVRGNSTGTLVWFFGNVDSGKIKHINFPPIIARIYIRLPHYSIRS 2160
Qy      1267  TLRMELMGCDLNSCSPMLGMSKAI SDAQITASSYFTNMPATWSPSKARLHLQGRSNWR 1326
Db      2161  TLRMELMGCDLNSCSPMLGMSKAI SDAQITASSYFTNMPATWSPSKARLHLQGRSNWR 2220
Qy      1327  POVNNPKEMLOVDFOKMTKVTGVTQGVKSLLTSMVYKEFLISSODGHOWTFFQNGKV 1386
Db      2221  POVNNPKEMLOVDFOKMTKVTGVTQGVKSLLTSMVYKEFLISSODGHOWTFFQNGKV 2280
Qy      1387  KVFQGNQDSFTPVVNSLDPILLTRYLRHPQSWHQAIALRMEVLGCEAQDLY 1438
Db      2281  KVFQGNQDSFTPVVNSLDPILLTRYLRHPQSWHQAIALRMEVLGCEAQDLY 2332

```

RESULT 15

```

PCT-US94-13200-2
; Sequence 2, Application PC/TUS9413200
; GENERAL INFORMATION:
; APPLICANT: Emory University
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 1100 Peachtree Street, Suite 2800
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/13200
; FILING DATE: 15-NOV-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Babst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: EMU106CIP(2)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-815-6508
; TELEFAX: 404-815-6555
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; TISSUE TYPE: Liver cDNA sequence
PCT-US94-13200-2

```

```

Query Match      94.1%; Score 7234; DB 5; Length 2332;
Best Local Similarity 61.7%; Pred. No. 0;
Matches 1438; Conservative 0; Mismatches 0; Indels 894; Gaps 1;

Qy      1  ATRRYILGAVELSWDMQSDLGELPVDARPPRPVKSPFPNTSVVYKTLFVEFTVHLFN 60
Db      1  ATRRYILGAVELSWDMQSDLGELPVDARPPRPVKSPFPNTSVVYKTLFVEFTVHLFN 60
Qy      61  IAKPRPFWMLGPTTQAAVYDVTVTITLKNMASHPVSLHAGVSYWKASGAEYDDQTSQ 120
Db      61  IAKPRPFWMLGPTTQAAVYDVTVTITLKNMASHPVSLHAGVSYWKASGAEYDDQTSQ 120
Qy      121  REKEDDKVPFGGSHTYVWQVLKENGPMASDPLCLITYSLSHVDLVKDLASGLIGALLVCR 180
Db      121  REKEDDKVPFGGSHTYVWQVLKENGPMASDPLCLITYSLSHVDLVKDLASGLIGALLVCR 180
Qy      181  EGS LAKEKTQTLHKFTILLFAVDEGKSMHSETKNSLMQDRDAASARAWPMTHTVNGVNR 240
Db      181  EGS LAKEKTQTLHKFTILLFAVDEGKSMHSETKNSLMQDRDAASARAWPMTHTVNGVNR 240
Qy      241  SUPGLIGCHRSYVHVIWGMGTTPVHSIFLGHTFLVNRHRCASLEISPIITLTAOTLL 300
Db      241  SUPGLIGCHRSYVHVIWGMGTTPVHSIFLGHTFLVNRHRCASLEISPIITLTAOTLL 300
Qy      301  MDLQGELLFCHISSHQHGMAYVKVDSCPEPQLRMKNNEAEYDDDLTISEMDVVR 360
Db      301  MDLQGELLFCHISSHQHGMAYVKVDSCPEPQLRMKNNEAEYDDDLTISEMDVVR 360
Qy      361  DDNSPSPFIQIRSVAKKHPTWVHYTAAEEDWDYAPLVLPDDRYSKSYQLNNGPQIRG 420

```







Db 121 REKEDKVPFGSGHYVWQVLKENGPMASDPLCLTYSHVLDLKDLSGLIGALLVCR 180  
Qy 181 EGSIAKEKQTTLHKETILLFAVDECKSWHSEFTKNSLMQDRDAASARAPKMHVNGYNR 240  
Db 181 EGSIAKEKQTTLHKETILLFAVDECKSWHSEFTKNSLMQDRDAASARAPKMHVNGYNR 240  
Qy 241 SLPLGLIGCHRSKVYWHVIGMGTTPEVHSIFLSEHTFLVRNHRQASLEISPIFTLTAQTLL 300  
Db 241 SLPLGLIGCHRSKVYWHVIGMGTTPEVHSIFLSEHTFLVRNHRQASLEISPIFTLTAQTLL 300  
Qy 301 MDLQGFLLFCHISSHQHDGMEAYVKVDCSPESPQLRMKNNEAEYDDDLTDTSEMDVVRP 360  
Db 301 MDLQGFLLFCHISSHQHDGMEAYVKVDCSPESPQLRMKNNEAEYDDDLTDTSEMDVVRP 360  
Qy 361 DDNSPSFIQIRSVAKKHPTWVHYIAAEEEDWDYAPLVLPDDRYSKYSLNNGPQIG 420  
Db 361 DDNSPSFIQIRSVAKKHPTWVHYIAAEEEDWDYAPLVLPDDRYSKYSLNNGPQIG 420  
Qy 421 RYKVKVRFMAYTDETFKTRAIQHSGILGPLLYGEVGDITLLIIFKNQASRPYNIYPHGI 480  
Db 421 RYKVKVRFMAYTDETFKTRAIQHSGILGPLLYGEVGDITLLIIFKNQASRPYNIYPHGI 480  
Qy 481 TDVRLYSRRLPKGVKHLKDFPILPGEIPKXKWTVTVEDGPKSDPRCLTRYISSFVNME 540  
Db 481 TDVRLYSRRLPKGVKHLKDFPILPGEIPKXKWTVTVEDGPKSDPRCLTRYISSFVNME 540  
Qy 541 RLASGLIGPLLI CYKESVDQGNQIMSDKRNVLFSVFDENRSWYLTENIORELPNPAG 600  
Db 541 RLASGLIGPLLI CYKESVDQGNQIMSDKRNVLFSVFDENRSWYLTENIORELPNPAG 600  
Qy 601 VQLEDEFOASNIMHSINGVYVDSLSQVCLHEVAYWYILSICAQTDPLSFVSGYTFKH 660  
Db 601 VQLEDEFOASNIMHSINGVYVDSLSQVCLHEVAYWYILSICAQTDPLSFVSGYTFKH 660  
Qy 661 KMVEDTLTLFPFSGTVMFMSMENPGLNLTGCHNSDFRNGMTALLKVSCKNTGDYFE 720  
Db 661 KMVEDTLTLFPFSGTVMFMSMENPGLNLTGCHNSDFRNGMTALLKVSCKNTGDYFE 720  
Qy 721 DSYEDISAYLLSKNNAIEPRSSONPPVLKREHOREITRTTLOSQDEIYDDTISVEMKK 780  
Db 721 DSYEDISAYLLSKNNAIEPRSSONPPVLKREHOREITRTTLOSQDEIYDDTISVEMKK 780  
Qy 781 EDPDIYDEBENQSPRFQKTRHYFLAAVERLMDYGMSSPHVLRNRAQSGVPQPKVV 840  
Db 781 EDPDIYDEBENQSPRFQKTRHYFLAAVERLMDYGMSSPHVLRNRAQSGVPQPKVV 840  
Qy 841 FQETDGSFTQPLRYGELNEHLGLGPYIRAEVEDNIMVTFRNQASRPYSFYSSLISYEE 900  
Db 841 FQETDGSFTQPLRYGELNEHLGLGPYIRAEVEDNIMVTFRNQASRPYSFYSSLISYEE 900  
Qy 901 DQOQAGPRKNFKVKNETKTYFWKVOHMAPTKDEEDCKAWAYPSDVLKDVHSGLIGP 960  
Db 901 DQOQAGPRKNFKVKNETKTYFWKVOHMAPTKDEEDCKAWAYPSDVLKDVHSGLIGP 960  
Qy 961 LLVCHTNTLNPAGRVQVTOEPALFTTIFDETSKWTFTENWERNCRAPCNIOQEDPTFXE 1020  
Db 961 LLVCHTNTLNPAGRVQVTOEPALFTTIFDETSKWTFTENWERNCRAPCNIOQEDPTFXE 1020  
Qy 1021 NYRFHAINGYIMDTLPGVMAQDQIRWYLLSGNSNENIHSIFSGHVFTVRKKEBYKMA 1080  
Db 1021 NYRFHAINGYIMDTLPGVMAQDQIRWYLLSGNSNENIHSIFSGHVFTVRKKEBYKMA 1080  
Qy 1081 LYNLYPGVFTVEMLPKAGIWRVECLIGELHAGMSTLFLVSNKQOTPLGMAASHIRD 1140  
Db 1081 LYNLYPGVFTVEMLPKAGIWRVECLIGELHAGMSTLFLVSNKQOTPLGMAASHIRD 1140  
Qy 1141 FOITASQYQWAPKPLARLYHSGSINAWSTKEPFWLKVLDLLAPMIHGIKTQAGRKFS 1200  
Db 1141 FOITASQYQWAPKPLARLYHSGSINAWSTKEPFWLKVLDLLAPMIHGIKTQAGRKFS 1200  
Qy 1201 SLXISQFIIMYSLDGKKWQYTRGNSTGTLVFFGNVDSSGIKHNIENPPIIARYLRHPT 1260  
Db 1201 SLXISQFIIMYSLDGKKWQYTRGNSTGTLVFFGNVDSSGIKHNIENPPIIARYLRHPT 1260

Qy 1261 HYSIRSTLRMELMGCDLNSCSPGLMESKASIDAQITASSYFTNNFATWSPSKARLHQ 1320  
Db 1261 HYSIRSTLRMELMGCDLNSCSPGLMESKASIDAQITASSYFTNNFATWSPSKARLHQ 1320  
Qy 1321 RSNARPOVNNPKWLQVDFOKTKMKVGTGTTQGVKSLTSMYVKEFLISSQDGHQWTLF 1380  
Db 1321 RSNARPOVNNPKWLQVDFOKTKMKVGTGTTQGVKSLTSMYVKEFLISSQDGHQWTLF 1380  
Qy 1381 FONGKVKVFCQGNQDSFTPVVNSLDPPLLTRYLRIHPQSWHQAIALRMEVLGCEAQDLY 1438  
Db 1381 FONGKVKVFCQGNQDSFTPVVNSLDPPLLTRYLRIHPQSWHQAIALRMEVLGCEAQDLY 1438

RESULT 2

US-10-047-257-1  
; Sequence 1, Application US/10047257  
; Publication No. US20020115152A1  
; GENERAL INFORMATION:  
; APPLICANT: Cho, Myung-Sam  
; APPLICANT: Chan, Sham-Yuen  
; APPLICANT: Kelsey, William  
; APPLICANT: Yee, Helena  
; TITLE OF INVENTION: Expression System for Factor VIII  
; FILE REFERENCE: MSB-7255.2  
; CURRENT APPLICATION NUMBER: US/10/047,257  
; CURRENT FILING DATE: 2002-01-15  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 1438  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Derived from  
; OTHER INFORMATION: human factor VIII sequence  
US-10-047-257-1

Query Match 100.0%; Score 7691; DB 14; Length 1438;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATRRYYLGAVELSWDYMOSDLGELPVDARPPRPVKSPFPNTSVVYKKTLFVEFTVHLFN 60  
Db 1 ATRRYYLGAVELSWDYMOSDLGELPVDARPPRPVKSPFPNTSVVYKKTLFVEFTVHLFN 60  
Qy 61 IAKPRPPMGLIGPTIOAEVYDVTVTILKNMASHVPSLHAGVSVYKASEGAEYDDQTSQ 120  
Db 61 IAKPRPPMGLIGPTIOAEVYDVTVTILKNMASHVPSLHAGVSVYKASEGAEYDDQTSQ 120  
Qy 121 REKEDKVPFGSGHYVWQVLKENGPMASDPLCLTYSHVLDLKDLSGLIGALLVCR 180  
Db 121 REKEDKVPFGSGHYVWQVLKENGPMASDPLCLTYSHVLDLKDLSGLIGALLVCR 180  
Qy 181 EGSIAKEKQTTLHKETILLFAVDECKSWHSEFTKNSLMQDRDAASARAPKMHVNGYNR 240  
Db 181 EGSIAKEKQTTLHKETILLFAVDECKSWHSEFTKNSLMQDRDAASARAPKMHVNGYNR 240  
Qy 241 SLPLGLIGCHRSKVYWHVIGMGTTPEVHSIFLSEHTFLVRNHRQASLEISPIFTLTAQTLL 300  
Db 241 SLPLGLIGCHRSKVYWHVIGMGTTPEVHSIFLSEHTFLVRNHRQASLEISPIFTLTAQTLL 300  
Qy 301 MDLQGFLLFCHISSHQHDGMEAYVKVDCSPESPQLRMKNNEAEYDDDLTDTSEMDVVRP 360  
Db 301 MDLQGFLLFCHISSHQHDGMEAYVKVDCSPESPQLRMKNNEAEYDDDLTDTSEMDVVRP 360  
Qy 361 DDNSPSFIQIRSVAKKHPTWVHYIAAEEEDWDYAPLVLPDDRYSKYSLNNGPQIG 420  
Db 361 DDNSPSFIQIRSVAKKHPTWVHYIAAEEEDWDYAPLVLPDDRYSKYSLNNGPQIG 420  
Qy 421 RYKVKVRFMAYTDETFKTRAIQHSGILGPLLYGEVGDITLLIIFKNQASRPYNIYPHGI 480  
Db 421 RYKVKVRFMAYTDETFKTRAIQHSGILGPLLYGEVGDITLLIIFKNQASRPYNIYPHGI 480

```
QY 481 TDVRLYSRRLPKGVKHLKDPFLLPGEIFKYKWTVTVEDGTPKSDPRCLTRYSSFNME 540
Db 481 TDVRLYSRRLPKGVKHLKDPFLLPGEIFKYKWTVTVEDGTPKSDPRCLTRYSSFNME 540
QY 541 RDLASGLIGPLLI CYKESVDQRGNQIMSDKRNVLFSVFDNRNRSWYLTENIQRLPNPAG 600
Db 541 RDLASGLIGPLLI CYKESVDQRGNQIMSDKRNVLFSVFDNRNRSWYLTENIQRLPNPAG 600
QY 601 VOLEDDPFOASNIHMSINGVYVDSLSQVCLHEVAYWYILSIGAQDTFLSVFFSGYTFKH 660
Db 601 VOLEDDPFOASNIHMSINGVYVDSLSQVCLHEVAYWYILSIGAQDTFLSVFFSGYTFKH 660
QY 661 KMVYEDTLTLFPFSGEIVFMSMENPGLWILGCHNSDFNRGWTALLKVSSCDKNTGYYE 720
Db 661 KMVYEDTLTLFPFSGEIVFMSMENPGLWILGCHNSDFNRGWTALLKVSSCDKNTGYYE 720
QY 721 DSYEDISAYLLSKNNAIBPRFSQNPVVKRQREITRTTLOSQBEIDYDDTISVEMKK 780
Db 721 DSYEDISAYLLSKNNAIBPRFSQNPVVKRQREITRTTLOSQBEIDYDDTISVEMKK 780
QY 781 EDFDIYDEDNQSPRSFOKTRHYFIAAVERLWDYGMSSSPHVLNRNRAQSGSVQFQKVV 840
Db 781 EDFDIYDEDNQSPRSFOKTRHYFIAAVERLWDYGMSSSPHVLNRNRAQSGSVQFQKVV 840
QY 841 FQETDGSFTQPLRGELNEHLGLGPYIRAEVEDNIMVTFRNQASRPYSFYSSLISYEE 900
Db 841 FQETDGSFTQPLRGELNEHLGLGPYIRAEVEDNIMVTFRNQASRPYSFYSSLISYEE 900
QY 901 DQOGABPRKQNFVFNPTKTYFWKQVHMAPTKDBFCKAWAYSDVDLEKDVHSLGILP 960
Db 901 DQOGABPRKQNFVFNPTKTYFWKQVHMAPTKDBFCKAWAYSDVDLEKDVHSLGILP 960
QY 961 LLVCHTNTLNPAHQROVTVQBFALFFTFIDETKSWYFTENMERNCRAPCNQIMEDPTPKE 1020
Db 961 LLVCHTNTLNPAHQROVTVQBFALFFTFIDETKSWYFTENMERNCRAPCNQIMEDPTPKE 1020
QY 1021 NYRPHAINGYIMDTLPGLVMAQDQRIWYLLSMGSNENIHSIHFGSHVFTVRKKEEYQMA 1080
Db 1021 NYRPHAINGYIMDTLPGLVMAQDQRIWYLLSMGSNENIHSIHFGSHVFTVRKKEEYQMA 1080
QY 1081 LYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTFLVYSNKCOPTPLGMAASHIRD 1140
Db 1081 LYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTFLVYSNKCOPTPLGMAASHIRD 1140
QY 1141 FOITASGQYGQWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMIHGIKTQGARQKPS 1200
Db 1141 FOITASGQYGQWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMIHGIKTQGARQKPS 1200
QY 1201 SLVYSQFTIMYSLDGKKWQTVRGNSGTGLMVFFGNVDSGSGIKHNI FNPPIIARYIRLHPT 1260
Db 1201 SLVYSQFTIMYSLDGKKWQTVRGNSGTGLMVFFGNVDSGSGIKHNI FNPPIIARYIRLHPT 1260
QY 1261 HYSIRSTRMELMGCDLNSCMPLGMSKASDAQITASSYFTNNFATWSPSKARLHIQ 1320
Db 1261 HYSIRSTRMELMGCDLNSCMPLGMSKASDAQITASSYFTNNFATWSPSKARLHIQ 1320
QY 1321 RSNARPOVNNPKWLOVDFOKTMKVTGVTQGVKSLTSMYKBEFLSSODGQHWTLF 1380
Db 1321 RSNARPOVNNPKWLOVDFOKTMKVTGVTQGVKSLTSMYKBEFLSSODGQHWTLF 1380
QY 1381 FQNGKVKVFGQNDQSFTPVNSLDPPLTRYLRHPQSVWVHQAIRMEVLGCEAQDLY 1438
Db 1381 FQNGKVKVFGQNDQSFTPVNSLDPPLTRYLRHPQSVWVHQAIRMEVLGCEAQDLY 1438
```

## RESULT 3

```
US-10-225-900-1
; Sequence 1, Application US/10225900
; Publication No. US2003007752A1
; GENERAL INFORMATION:
; APPLICANT: Cho, Myung-Sam
; APPLICANT: Chan, Sham-Yuen
```

```
; APPLICANT: Kelsey, William
; APPLICANT: Yee, Helena
; TITLE OF INVENTION: Expression System for Factor VIII
; FILE REFERENCE: MSB-7255
; CURRENT APPLICATION NUMBER: US/10/225,900
; CURRENT FILING DATE: 2002-08-22
; PRIOR APPLICATION NUMBER: US/09/209,916
; PRIOR FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1438
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Derived from
; OTHER INFORMATION: human factor VIII sequence
US-10-225-900-1
```

```
Query Match 100.0%; Score 7691; DB 15; Length 1438;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATRRYYLGAVELSWDMQSDLGELPVDARPPRPVKSPFPNTSVYKTLFVEFTVHLFN 60
Db 1 ATRRYYLGAVELSWDMQSDLGELPVDARPPRPVKSPFPNTSVYKTLFVEFTVHLFN 60
QY 61 IAKPRPPWMLGLGPTIQAEVYDVITLKNWASHPVSLHAGVSYWKASEGAEYDDQTSQ 120
Db 61 IAKPRPPWMLGLGPTIQAEVYDVITLKNWASHPVSLHAGVSYWKASEGAEYDDQTSQ 120
QY 121 REKEDDKVFFGSGHYVQVQLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 180
Db 121 REKEDDKVFFGSGHYVQVQLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 180
QY 181 EGSIAKKTQTLHKFILLFAVFDGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNR 240
Db 181 EGSIAKKTQTLHKFILLFAVFDGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNR 240
QY 241 SLPLGICHRKSVVHWVIGMGTTPDEVHSIFLEGHTFLVRNHRQASLEISPIITFLTAOTLL 300
Db 241 SLPLGICHRKSVVHWVIGMGTTPDEVHSIFLEGHTFLVRNHRQASLEISPIITFLTAOTLL 300
QY 301 MDLQOQLLFCHISSHQHDGMEAYVKVDSCPEEPQLRMKNNEEADYDDDLTDSMDVVRFF 360
Db 301 MDLQOQLLFCHISSHQHDGMEAYVKVDSCPEEPQLRMKNNEEADYDDDLTDSMDVVRFF 360
QY 361 DDNSPSFIQIRSVAKGHPKWTWVHYIAAEEDWDYAPLVLPADDDRSYKSYQVNNGPQIRG 420
Db 361 DDNSPSFIQIRSVAKGHPKWTWVHYIAAEEDWDYAPLVLPADDDRSYKSYQVNNGPQIRG 420
QY 421 RYKVKVRFMAYTDEFTKTRRAIOHESGILGPLLYGEVGDITLLIIFKNQASRPNTYPHGI 480
Db 421 RYKVKVRFMAYTDEFTKTRRAIOHESGILGPLLYGEVGDITLLIIFKNQASRPNTYPHGI 480
QY 481 TDVRLYSRRLPKGVKHLKDPFLLPGEIFKYKWTVTVEDGTPKSDPRCLTRYSSFNME 540
Db 481 TDVRLYSRRLPKGVKHLKDPFLLPGEIFKYKWTVTVEDGTPKSDPRCLTRYSSFNME 540
QY 541 RDLASGLIGPLLI CYKESVDQRGNQIMSDKRNVLFSVFDNRNRSWYLTENIQRLPNPAG 600
Db 541 RDLASGLIGPLLI CYKESVDQRGNQIMSDKRNVLFSVFDNRNRSWYLTENIQRLPNPAG 600
QY 601 VOLEDDPFOASNIHMSINGVYVDSLSQVCLHEVAYWYILSIGAQDTFLSVFFSGYTFKH 660
Db 601 VOLEDDPFOASNIHMSINGVYVDSLSQVCLHEVAYWYILSIGAQDTFLSVFFSGYTFKH 660
QY 661 KMVYEDTLTLFPFSGEIVFMSMENPGLWILGCHNSDFNRGWTALLKVSSCDKNTGYYE 720
Db 661 KMVYEDTLTLFPFSGEIVFMSMENPGLWILGCHNSDFNRGWTALLKVSSCDKNTGYYE 720
QY 721 DSYEDISAYLLSKNNAIBPRFSQNPVVKRQREITRTTLOSQBEIDYDDTISVEMKK 780
Db 721 DSYEDISAYLLSKNNAIBPRFSQNPVVKRQREITRTTLOSQBEIDYDDTISVEMKK 780
```

```
Db 721 DSYEDISAYLLSKNNAIEPRSFQNPVVKRQREITRTTLLQSDQEEIDYDDTISVEMKK 780
Qy 781 EDYDIDYDENQSPRSFOKTRHYFIAAVERLWDYGMSSSPHVLNRQAQSGSVQFKKVV 840
Db 781 EDYDIDYDENQSPRSFOKTRHYFIAAVERLWDYGMSSSPHVLNRQAQSGSVQFKKVV 840
Qy 841 FQEFDTGSGFTQPLRGELNEHLGLPGYIRAEVDNIMVTFRNQASRPYSYSSLSIYEE 900
Db 841 FQEFDTGSGFTQPLRGELNEHLGLPGYIRAEVDNIMVTFRNQASRPYSYSSLSIYEE 900
Qy 901 DQOQAEPRKQFVKENETKTFWKVQHMAFKDBFCKKAWAYPSDVLDKDVHSGILGP 960
Db 901 DQOQAEPRKQFVKENETKTFWKVQHMAFKDBFCKKAWAYPSDVLDKDVHSGILGP 960
Qy 961 LLVCHTNTLNPAHQGVQVTFQSFALPFTTFDETKSWYFTENMERNCRACNTQMEDPTKE 1020
Db 961 LLVCHTNTLNPAHQGVQVTFQSFALPFTTFDETKSWYFTENMERNCRACNTQMEDPTKE 1020
Qy 1021 NYRFAINGYIMDTLPGLVMAQDQRIRWYLLSMGNSNENIHSIHFGSHVFTVRKKEEYKWA 1080
Db 1021 NYRFAINGYIMDTLPGLVMAQDQRIRWYLLSMGNSNENIHSIHFGSHVFTVRKKEEYKWA 1080
Qy 1081 LYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPLGMASGHIRD 1140
Db 1081 LYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPLGMASGHIRD 1140
Qy 1141 FOITASGOYQWAPKLARLHYSGSINAWSTKBPFSWIKVDLLAPMIHGKIKTOGARQKFS 1200
Db 1141 FOITASGOYQWAPKLARLHYSGSINAWSTKBPFSWIKVDLLAPMIHGKIKTOGARQKFS 1200
Qy 1201 SLIYISQFIIMYSLDGKQWQYRGNSTGTLMWFFGNVSSGIKHNIFFNPPIIARYIRLHPT 1260
Db 1201 SLIYISQFIIMYSLDGKQWQYRGNSTGTLMWFFGNVSSGIKHNIFFNPPIIARYIRLHPT 1260
Qy 1261 HYSIRSTRLMELMGCDLNSCSPMLGSKAISDAQITASSYFTNMFATWSPSKARLHLQ 1320
Db 1261 HYSIRSTRLMELMGCDLNSCSPMLGSKAISDAQITASSYFTNMFATWSPSKARLHLQ 1320
Qy 1321 RSNARWQVNNPKEWLQVDFQTKMTVGTTOGVKSLLTSMVVKGFLLSSSDGQHWTLF 1380
Db 1321 RSNARWQVNNPKEWLQVDFQTKMTVGTTOGVKSLLTSMVVKGFLLSSSDGQHWTLF 1380
Qy 1381 FQNGKVKVQGNQDSFTPVVNSLDPPLLRILRHPQSWHQAIALRMEVLGCEAQDLY 1438
Db 1381 FQNGKVKVQGNQDSFTPVVNSLDPPLLRILRHPQSWHQAIALRMEVLGCEAQDLY 1438
```

## RESULT 4

```
US-10-095-718-2
; Sequence 2, Application US/10095718
; Publication No. US20020131956A1
; GENERAL INFORMATION:
; APPLICANT: Walsh, Christopher
; APPLICANT: Chao, Hengjun
; APPLICANT: Burstein, Haim
; APPLICANT: Lynch, Carmel
; APPLICANT: Stepan, Tony
; APPLICANT: Munson, Keith
; TITLE OF INVENTION: Adeno-Associated Virus Vectors Encoding Factor VIII and
; TITLE OF INVENTION: Methods of Using the Same
; FILE REFERENCE: 35052/204375
; CURRENT APPLICATION NUMBER: US/10/095,718
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 09/689,430
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/158,780
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1471
; TYPE: PRT
; ORGANISM: Homo sapiens B-domain deleted factor VIII
```

```
; FEATURE:
; OTHER INFORMATION: Homo sapiens BDD FVIII
US-10-095-718-2

Query Match 99.8%; Score 7674; DB 14; Length 1471;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1436; Conservative 0; Mismatches 0; Indels 14; Gaps 1;

Qy 1 ATRRYILGAVELSDWYMQSDLGELPVDARPPRPVKSPFPNTSVYVKTTLFVEPTVHLFN 60
Db 20 ATRRYILGAVELSDWYMQSDLGELPVDARPPRPVKSPFPNTSVYVKTTLFVEPTVHLFN 79
Qy 61 IAKRPPWMLGILGTIOAEVYDVTWITLKNMASHPVLSHAGVSYWKAEGAEVDDQTSQ 120
Db 80 IAKRPPWMLGILGTIOAEVYDVTWITLKNMASHPVLSHAGVSYWKAEGAEVDDQTSQ 139
Qy 121 REKEDDKVFPQSGSHYVQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 180
Db 140 REKEDDKVFPQSGSHYVQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 199
Qy 181 EGSLAKEKTQTLHKFILLFAVDFDEGKSWHSETKNSLMQDRDAASARAWPKMHTVGVYNR 240
Db 200 EGSLAKEKTQTLHKFILLFAVDFDEGKSWHSETKNSLMQDRDAASARAWPKMHTVGVYNR 259
Qy 241 SLPLGIGCHRSVYVHWIGMGTTFEVHSIFLEGHTFLVRNHRQASLEISPIITLTAQTLL 300
Db 260 SLPLGIGCHRSVYVHWIGMGTTFEVHSIFLEGHTFLVRNHRQASLEISPIITLTAQTLL 319
Qy 301 MDLQGFLLFCHISHOHDGMEAYVKVDSCPEPOLRMKNNEEAEDYDDDLTDSMDVYVRF 360
Db 320 MDLQGFLLFCHISHOHDGMEAYVKVDSCPEPOLRMKNNEEAEDYDDDLTDSMDVYVRF 379
Qy 361 DDDNSPSFIQIRSVAKKHPKTVVHYIAAEEDWDYAPLVLAPDPRRSYKQYLNNQPORG 420
Db 380 DDDNSPSFIQIRSVAKKHPKTVVHYIAAEEDWDYAPLVLAPDPRRSYKQYLNNQPORG 439
Qy 421 RYKVKRPMAYTDEFTKREAIQHSGLGLGELLGEGVDTLLIIFKNQASPNYIYPGI 480
Db 440 RYKVKRPMAYTDEFTKREAIQHSGLGLGELLGEGVDTLLIIFKNQASPNYIYPGI 499
Qy 481 TDVRLYRRRLPKGVKHLKDPFLLPGEIFKTKWTVVBDGPTKSDPRCLTYYSFVNME 540
Db 500 TDVRLYRRRLPKGVKHLKDPFLLPGEIFKTKWTVVBDGPTKSDPRCLTYYSFVNME 559
Qy 541 RDLASGLIGPLLCYKESVDQGNQIMSDKRNVLFSVFDENRSLYLTENIQRFPLNPAG 600
Db 560 RDLASGLIGPLLCYKESVDQGNQIMSDKRNVLFSVFDENRSLYLTENIQRFPLNPAG 619
Qy 601 VQLEDPEFQASNMHISINGYVFDLSQLSVCLHEVAYWYILSGACTDLSVFFSGYTRKH 660
Db 620 VQLEDPEFQASNMHISINGYVFDLSQLSVCLHEVAYWYILSGACTDLSVFFSGYTRKH 679
Qy 661 KMVYEDTTLTPPFSGETVFMGMENPGLMILGCHNSDFNRGMTALLKYSSCDKNTGDIYE 720
Db 680 KMVYEDTTLTPPFSGETVFMGMENPGLMILGCHNSDFNRGMTALLKYSSCDKNTGDIYE 739
Qy 721 DSYEDISAYLLSKNNAIEPRSFQNPVVKRQREITRTTLLQSDQEEIDYDDTISVEMKK 766
Db 740 DSYEDISAYLLSKNNAIEPRSFQNPVVKRQREITRTTLLQSDQEEIDYDDTISVEMKK 799
Qy 767 EIDYDDTISVEMKKEDFDIYDENQSPRSFOKTRHYFIAAVERLWDYGMSSSPHVLNR 826
Db 800 EIDYDDTISVEMKKEDFDIYDENQSPRSFOKTRHYFIAAVERLWDYGMSSSPHVLNR 859
Qy 827 RAQSGSVQFKKVVQEFDTGSGFTQPLRGELNEHLGLPGYIRAEVDNIMVTFRNQAS 886
Db 860 RAQSGSVQFKKVVQEFDTGSGFTQPLRGELNEHLGLPGYIRAEVDNIMVTFRNQAS 919
Qy 887 REYSFYSSLSIYEEQROGAEPKRNFKVFNENETKTYFWKQVQHMAFKDBFCKKAWAYPSD 946
Db 920 REYSFYSSLSIYEEQROGAEPKRNFKVFNENETKTYFWKQVQHMAFKDBFCKKAWAYPSD 979
Qy 947 VDLEKDVHSGILGLPGLLVCHTNTLNPAHQGVQVTFQSFALPFTTFDETKSWYFTENMERNCR 1006
```

Db 980 VLEKDVHSLGILLVCHTNTLNPAGHQVTVQSFALFTTIDETKSYFTENMERNCR 1039  
QY 1007 APCNIQMEPTFKENYRFHAINGYIMDTPLGLVMAQDQIRWYLLSGNSNENIHSYHFSG 1066  
Db 1040 APCNIQMEPTFKENYRFHAINGYIMDTPLGLVMAQDQIRWYLLSGNSNENIHSYHFSG 1099  
QY 1067 HVFTVRKKEBYKALYNLYPGVPETVEMLPSPKAGIWRVECLIGEHLHAGMSTLFLVYSNK 1126  
Db 1100 HVFTVRKKEBYKALYNLYPGVPETVEMLPSPKAGIWRVECLIGEHLHAGMSTLFLVYSNK 1159  
QY 1127 CQTPLGWASHIIRDFQITASQYQGWAPKARLHYSGSINAWSTKPPFWIKVDLLAPMI 1186  
Db 1160 CQTPLGWASHIIRDFQITASQYQGWAPKARLHYSGSINAWSTKPPFWIKVDLLAPMI 1219  
QY 1187 IHGKITQAROKFSSLYISQFIIMYSLDGKKWOTYRGNSTGLTMVFFGNVDSGGIKHNIF 1246  
Db 1220 IHGKITQAROKFSSLYISQFIIMYSLDGKKWOTYRGNSTGLTMVFFGNVDSGGIKHNIF 1279  
QY 1247 NPPIIARYIRLPHTHYSIRSLRMLWGCGLNSCSMPLGMSKAISSDAQITASSYFTNMW 1306  
Db 1280 NPPIIARYIRLPHTHYSIRSLRMLWGCGLNSCSMPLGMSKAISSDAQITASSYFTNMW 1339  
QY 1307 ATWSPSKARLHLQGRSNAMRPQVNNPKEMLOVDFOKTMKVTGVTTOGVKSLLTSMYKBF 1366  
Db 1340 ATWSPSKARLHLQGRSNAMRPQVNNPKEMLOVDFOKTMKVTGVTTOGVKSLLTSMYKBF 1399  
QY 1367 L1SSSDQGHQWTLFFQNGKVKVQGNQDSFTPVVNSLDPPLLTRYLRIRHPQSWHQAIALR 1426  
Db 1400 L1SSSDQGHQWTLFFQNGKVKVQGNQDSFTPVVNSLDPPLLTRYLRIRHPQSWHQAIALR 1459  
QY 1427 MEVLGCEAODLY 1438  
Db 1460 MEVLGCEAODLY 1471

## RESULT 5

US-09-957-641-2  
; Sequence 2, Application US/09957641  
; Publication No. US20020182670A1  
; GENERAL INFORMATION:  
; APPLICANT: Emory University  
; TITLE OF INVENTION: MODIFIED FACTOR VIII  
; FILE REFERENCE: 75-00  
; CURRENT APPLICATION NUMBER: US/09/957,641  
; CURRENT FILING DATE: 2001-09-16  
; PRIOR APPLICATION NUMBER: US 60/234047  
; PRIOR FILING DATE: 2000-09-19  
; PRIOR APPLICATION NUMBER: US 60/236460  
; PRIOR FILING DATE: 2000-09-29  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 2332  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-957-641-2

Query Match 94.1%; Score 7234; DB 10; Length 2332;

Best Local Similarity 61.7%; Pred. No. 0;

Matches 1438; Conservative 0; Mismatches 0; Indels 894; Gaps 1;

QY 1 ATRRYIAGVLSWDYMQSDGLPVDARPPRPVKSPFPNTSVVYKKTLPVEFTVHLEN 60  
Db 1 ATRRYIAGVLSWDYMQSDGLPVDARPPRPVKSPFPNTSVVYKKTLPVEFTVHLEN 60  
QY 61 IAKPRPWWGLLQPTQAEYVDYVITLKNWASHPVSLHAGVSYWKASGAYDDQTSQ 120  
Db 61 IAKPRPWWGLLQPTQAEYVDYVITLKNWASHPVSLHAGVSYWKASGAYDDQTSQ 120  
QY 121 REKEDKVPFGSGHTYVQVLKNGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 180  
Db 121 REKEDKVPFGSGHTYVQVLKNGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 180

QY 181 EGSLAKBKOTTLHKFTLLFAVFDGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNR 240  
Db 181 EGSLAKBKOTTLHKFTLLFAVFDGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNR 240  
QY 241 SLPGILGCHRKSVYHVICMGTTPEVHSIFLGHFTFLVNRHQASLEISPIITFLTAQTLL 300  
Db 241 SLPGILGCHRKSVYHVICMGTTPEVHSIFLGHFTFLVNRHQASLEISPIITFLTAQTLL 300  
QY 301 MDLQGFLLFCHTSSHQDQMEAYVKVDSCEPQPLRMKNNEEAEDYDDDLTSEMDVVF 360  
Db 301 MDLQGFLLFCHTSSHQDQMEAYVKVDSCEPQPLRMKNNEEAEDYDDDLTSEMDVVF 360  
QY 361 DDNSPSFIQIRSVAKKHPTKTVHYIAAEEEDWDYAPLVLAPDDRYSKYQYLNNQPQIG 420  
Db 361 DDNSPSFIQIRSVAKKHPTKTVHYIAAEEEDWDYAPLVLAPDDRYSKYQYLNNQPQIG 420  
QY 421 RYKVKRVMAYTDETFKTRTREALQHSGLIGPLLYGEGVGTLLIIFKNQASRPNIYPHGI 480  
Db 421 RYKVKRVMAYTDETFKTRTREALQHSGLIGPLLYGEGVGTLLIIFKNQASRPNIYPHGI 480  
QY 481 TVRPLYSRRLPGVVKHLKDFPILPGEIFKYKWTVTVEDGPTKSDPRCLTRYYSFVNMW 540  
Db 481 TVRPLYSRRLPGVVKHLKDFPILPGEIFKYKWTVTVEDGPTKSDPRCLTRYYSFVNMW 540  
QY 541 RDLASGLIGPLLYCYKESVDQRGNQIMSDKRNVLFSVDFENRSWYLTENIQRFLENPAG 600  
Db 541 RDLASGLIGPLLYCYKESVDQRGNQIMSDKRNVLFSVDFENRSWYLTENIQRFLENPAG 600  
QY 601 VQLEDPFOASNIHMSINGYVFDLSQLSVCLHEVAYWYILSQAQDTFISVFPFGTTFKH 660  
Db 601 VQLEDPFOASNIHMSINGYVFDLSQLSVCLHEVAYWYILSQAQDTFISVFPFGTTFKH 660  
QY 661 KMVYEDTLTLPFSGETVFMENPGLMILGCHNSDFRNKGMTALLKVSCKNTGDYIE 720  
Db 661 KMVYEDTLTLPFSGETVFMENPGLMILGCHNSDFRNKGMTALLKVSCKNTGDYIE 720  
QY 721 DSYEDISAYLLSKNNAIEPRSF----- 742  
Db 721 DSYEDISAYLLSKNNAIEPRSFQNSRHPSTRQKQFNATTIPENDIEKTDPMFAHRTMP 780  
QY 743 ----- 742  
Db 781 KIQNVSSDLLMLLRQSPHGLSLDLQEAQYETFPDOPSPGAIDNNNSLSEMTFRPQ 840  
QY 743 ----- 742  
Db 841 LHHSGDMVTPPSGLQLRLNEKLGTTAATELKKLDPKVSTSNLSTIPSDNLAAGTDN 900  
QY 743 ----- 742  
Db 901 TSSLGPPMPVHYDSQDITLFGKKSSPLTESGGPLSLSENNDSKLLESGLMNSQESSW 960  
QY 743 ----- 742  
Db 961 GKNVSTESGRFLFKGRAHPALLTKDNALFKVSISSLKTKNTKNSATNRKTHIDGPSL 1020  
QY 743 ----- 742  
Db 1021 LIENSPPVQNTLESDETFKVTPLIHDRMLMDKNATALLNHNMSKNTSSKNMVMVQQ 1080  
QY 743 ----- 742  
Db 1081 KEGPIPPDAQNPMDSFFKMLFLPESARWIQRTGKNLSNGQSPSPKQLVSLGPEKSVEG 1140  
QY 743 ----- 742  
Db 1141 QNFLEKKNVVKGSGFTKDVGLKEMVFPSSRNLFLTNLDNLHNHNTHNQEKIQEEIEK 1200  
QY 743 ----- 742  
Db 1201 KETLIOENVLPQIHTVTGTGTFKMKNLFLLSSTRQNVESYGAYAPVLQDFRSLDSTNR 1260

Qy 743 ----- 742  
Db 1261 TKHTAHFSKKGBENLEGLNQTQKQIVEKYACTTRISPNTSQNFVTVQSRKALKQFRL 1320  
Qy 743 ----- 742  
Db 1321 PLEETELEKRIIVDDTSTQSKNMKHLTPSTLTQIDYNEKEKGAITQSPGSDCLTRSHSI 1380  
Qy 743 ----- 742  
Db 1381 PQANRSPPIAKVSPFPIRPIYLRVLFDQNSHLPAASYRKXKDSGVQBSHFLQAKK 1440  
Qy 743 ----- 742  
Db 1441 NNLSLAULTLEMTGDQREVGSIGTSATNSVTVYKVKVENTVLKPKPOLPKTSGKVELLPKVIH 1500  
Qy 743 ----- 742  
Db 1501 YOKOLFPPTETNGSPGHLDLVEGSLLOQTEGAIKWNEANRFGKVPFLRVATESSAKTPSK 1560  
Qy 743 ----- 742  
Db 1561 LLDPLANDHNYGTQPKBEWKSQESPEKTAFKKDDTILSNACSNHAIANEGQNK 1620  
Qy 743 ----- 742  
Db 1621 EIEVTWAKQRTERLCSQNPVVKHQREITRTTLOSQDEEIDYDDTISVENKKEDFDIY 1680  
Qy 787 DEDENQSPRSQKTRHYFIAAVERLDYGMSSPHVLRNRAQSGVPOPKVVOEFTD 846  
Db 1681 DEDENQSPRSQKTRHYFIAAVERLDYGMSSPHVLRNRAQSGVPOPKVVOEFTD 1740  
Qy 847 GSFTQPLVGEHLGGLGPVIRAEVEDNIMVTFRNQASRPSYSSILSYEEEDQOQA 906  
Db 1741 GSFTQPLVGEHLGGLGPVIRAEVEDNIMVTFRNQASRPSYSSILSYEEEDQOQA 1800  
Qy 907 EPRKNFVKNETKTYFWKVOHMAPTKQBFDCAKAWYSDVDLEKDVHSGILGILLVCHT 966  
Db 1801 EPRKNFVKNETKTYFWKVOHMAPTKQBFDCAKAWYSDVDLEKDVHSGILGILLVCHT 1860  
Qy 967 NTLNPAHQVTVQSFALFTTIFDETKSWYFTENMERNCRACNTQMEDPTFKENYRPHA 1026  
Db 1861 NTLNPAHQVTVQSFALFTTIFDETKSWYFTENMERNCRACNTQMEDPTFKENYRPHA 1920  
Qy 1027 INGYIMDTLPGVMAQDQIRWYLLSMGSENIHSHFSGHVTFRKKEEYKALYNLYP 1086  
Db 1921 INGYIMDTLPGVMAQDQIRWYLLSMGSENIHSHFSGHVTFRKKEEYKALYNLYP 1980  
Qy 1087 GVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPGMAHGHRDFOITAS 1146  
Db 1981 GVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPGMAHGHRDFOITAS 2040  
Qy 1147 GOYGOWAPKLARLHYSGSNAMSTKEPFWIKVDLLAPMIHGIKTQGARQFSSLYISQ 1206  
Db 2041 GOYGOWAPKLARLHYSGSNAMSTKEPFWIKVDLLAPMIHGIKTQGARQFSSLYISQ 2100  
Qy 1207 FIIMYSLDGKKWQTVRGNSTGTLMVFFGNVDSGKHNIFNPPIIARYIRLHPHYISRS 1266  
Db 2101 FIIMYSLDGKKWQTVRGNSTGTLMVFFGNVDSGKHNIFNPPIIARYIRLHPHYISRS 2160  
Qy 1267 TLRMELMGCDLNSCSMPLGMESKAISDAQITASSYFTNPMATWSPSKARLHLQGRNMR 1326  
Db 2161 TLRMELMGCDLNSCSMPLGMESKAISDAQITASSYFTNPMATWSPSKARLHLQGRNMR 2220  
Qy 1327 PQVNNKEMLOVDFOKTMKVTGTTQGVKSLLTSMYVKEFLISSODGHQWTLFPONGKV 1386  
Db 2221 PQVNNKEMLOVDFOKTMKVTGTTQGVKSLLTSMYVKEFLISSODGHQWTLFPONGKV 2280  
Qy 1387 KVFQGNQDSFTPVVNSLDPLLTLYLRHPQSWHQAIRMEVYLGEAODLY 1438  
Db 2281 KVFQGNQDSFTPVVNSLDPLLTLYLRHPQSWHQAIRMEVYLGEAODLY 2332

## RESULT 6

US-10-131-510A-2  
; Sequence 2, Application US/10131510A  
; Publication No. US20030166536A1  
; GENERAL INFORMATION:  
; APPLICANT: Lollar, John S  
; TITLE OF INVENTION: Modified Factor VIII  
; FILE REFERENCE: 75-95J  
; CURRENT APPLICATION NUMBER: US/10/131,510A  
; CURRENT FILING DATE: 2002-07-10  
; PRIOR APPLICATION NUMBER: U.S. 09/315,179  
; PRIOR FILING DATE: 1999-05-20  
; PRIOR APPLICATION NUMBER: U.S. 09/037,601  
; PRIOR FILING DATE: 1998-03-10  
; PRIOR APPLICATION NUMBER: U.S. 08/670,707  
; PRIOR FILING DATE: 1996-06-26  
; PRIOR APPLICATION NUMBER: PCT/US97/11155  
; PRIOR FILING DATE: 1997-06-26  
; PRIOR APPLICATION NUMBER: PCT/US94/13200  
; PRIOR FILING DATE: 1994-11-15  
; PRIOR APPLICATION NUMBER: U.S. 08/212,133  
; PRIOR FILING DATE: 1994-03-11  
; PRIOR APPLICATION NUMBER: U.S. 07/864,004  
; PRIOR FILING DATE: 1992-04-07  
; NUMBER OF SEQ ID NOS: 40  
; SOFTWARE: Patent in ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 2332  
; TYPE: PNT  
; ORGANISM: Homo sapiens  
US-10-131-510A-2

Query Match 94.1%; Score 7234; DB 12; Length 2332;

Best Local Similarity 61.7%; Pred. No. 0;

Matches 1438; Conservative 0; Mismatches 0; Indels 894; Gaps 1;

Qy 1 ATRRYTLCAVELSDWYMSQDLGELPVDARFPVPKSPPTNTSVVYKKTLPVEFTVHLFN 60  
Db 1 ATRRYTLCAVELSDWYMSQDLGELPVDARFPVPKSPPTNTSVVYKKTLPVEFTVHLFN 60  
Qy 61 IAKRPPPMWGLGPTTQAEVYDVTVTILKNMASHPVSLHVGVSYWKASGAEYDDQTSQ 120  
Db 61 IAKRPPPMWGLGPTTQAEVYDVTVTILKNMASHPVSLHVGVSYWKASGAEYDDQTSQ 120  
Qy 121 REKEDDKVPFGSGHTYVQVLKENGPMASDPLCLTYSYLVSHVDLVKDLNSGLIGALLVCR 180  
Db 121 REKEDDKVPFGSGHTYVQVLKENGPMASDPLCLTYSYLVSHVDLVKDLNSGLIGALLVCR 180  
Qy 181 EGSLAKEKTQTLHKFILLFAVFDGKSWHSETKNSLMQDRDAASARAMPKMHVTNGYVNR 240  
Db 181 EGSLAKEKTQTLHKFILLFAVFDGKSWHSETKNSLMQDRDAASARAMPKMHVTNGYVNR 240  
Qy 241 SLPLGLIGCHRSVYWHVIGMGTTPRVHSIFLEGHTFLVNRHQASLEISPTIFLTAQTLL 300  
Db 241 SLPLGLIGCHRSVYWHVIGMGTTPRVHSIFLEGHTFLVNRHQASLEISPTIFLTAQTLL 300  
Qy 301 MDLGGQFLLFCHISSHODGMEAYKVDSCPEPQLRMKNNEAEYDDDLTDSEMDVVRP 360  
Db 301 MDLGGQFLLFCHISSHODGMEAYKVDSCPEPQLRMKNNEAEYDDDLTDSEMDVVRP 360  
Qy 361 DDNSPSPFIQIRSVAKKHPKTHVYIAAEEEDWDYAPLVLPDDRYSKSOYLNNGPQIRG 420  
Db 361 DDNSPSPFIQIRSVAKKHPKTHVYIAAEEEDWDYAPLVLPDDRYSKSOYLNNGPQIRG 420  
Qy 421 RYKVKVRFWAYTDETFKTRREALQHSGLIGPLLYGEVGDITLIIIFKNQASRPNTYPHGI 480  
Db 421 RYKVKVRFWAYTDETFKTRREALQHSGLIGPLLYGEVGDITLIIIFKNQASRPNTYPHGI 480  
Qy 481 TDVRPLYSRRLPKGVKHLKDFPILPGEIPKFKWTVTVEDGPTKSDPRCLTRYSSVFNME 540  
Db 481 TDVRPLYSRRLPKGVKHLKDFPILPGEIPKFKWTVTVEDGPTKSDPRCLTRYSSVFNME 540  
Qy 541 RDLASGLIGPLLIICYKESVDQRGNQIMSDKRVNVLFSVFDENRSWLTENIQRFLEPNPAG 600

Db 541 RDLASGLIGLLICYSVDRGQINSDKRNVLFSVFDENRSWYLTENIRFLPNPAG 600  
Qy 601 VQLEDPFOASINMHSINGVYVDSLQLSVCLHEVAYWYLLSIGAQDTFLSVFFSGYTFKH 660  
Db 601 VQLEDPFOASINMHSINGVYVDSLQLSVCLHEVAYWYLLSIGAQDTFLSVFFSGYTFKH 660  
Qy 661 KMVEDTLTLFPFSGEYVFMSEMPGLWILGCHNSDFRNGMTALLKVSSCDKNTGDIYE 720  
Db 661 KMVEDTLTLFPFSGEYVFMSEMPGLWILGCHNSDFRNGMTALLKVSSCDKNTGDIYE 720  
Qy 721 DSYEDISAYLLSKNNALEPRSF 742  
Db 721 DSYEDISAYLLSKNNALEPRSFQNSRHPSTROKQFNATTIPENDIKTDPFWAHRTMP 780  
Qy 743 742  
Db 781 KIQNVSSDLLMLLRQSPTHGLSLSDLOQAEYETFSDDPSGALDSNNSISEMTHFRPQ 840  
Qy 743 742  
Db 841 LHHSGDMVFTPESGLOLRLEKLGTTAATBELKKLDKFKVST'SNNLISTIPSDNLAAGTDN 900  
Qy 743 742  
Db 901 TSSLGPPSPVHVDSQDSTTLFGKKSPLTESGGPLSLSEENNDKLLSGLMNSQSSW 960  
Qy 743 742  
Db 961 GKNVSTESGRLFKGKAHGPALLTKONALFKVSIILLKTKTNSNNSATNRKTHIDGPSL 1020  
Qy 743 742  
Db 1021 LIENSPSWQWNILESDETFKVKVTPLIHDMMLMDKNATALRLNHSNNTTSSKNMEMYQOK 1080  
Qy 743 742  
Db 1081 KEGFIPDPAQNDMSFFKMLFLPESARWIQTHGKNSLNSGGQSPKQLVSLGPEKSVEG 1140  
Qy 743 742  
Db 1141 QNFLSEKNVVGGEFTKDVGLKEMVFPSSRNLFNLNHLNHNTHNQEKIQQEIEK 1200  
Qy 743 742  
Db 1201 KETLIOENVVLPOIHTVTGTKNFMKNLFLSTRQNVESYEGAYAPVLQDFRSLNDSTNR 1260  
Qy 743 742  
Db 1261 TKKHTAHFSKGBEENLEGLNQTQIWEKYACTTRISPNTSQNFVTVQSRKALKQFRL 1320  
Qy 743 742  
Db 1321 PLBETELEKRIIVDDTSTQSKMKHUTPSTLTQIDYNEKEGAIQSPSLDCILTRSHSI 1380  
Qy 743 742  
Db 1381 PQANRSPLIAKVSSPFSIRPIYLRVLFDQNSHLPASVRKDSGVQSSSHFLQAGK 1440  
Qy 743 742  
Db 1441 NNLSLAILTLEMTGDQREVSLGTSATNSVTYKKVENTVLPKPLDKPTSGKVELLPKVI 1500  
Qy 743 742  
Db 1501 YQKDLPTTETSNGSPGHLDIVESLLOGTEGAIKNANRPGKVPFLRVATESAKTPSK 1560  
Qy 743 742  
Db 1561 LLDPLANDNHYGTQIPKEEWSQESPEKTAFFKKOTILSNACENHATAINEGONKP 1620  
Qy 743 -----SQNPVLRKHOREITRTTLOSDBEIDYDDTISVEMKEDFDIY 786  
|||||

Db 1621 EIEVTWAKQRTRELCSQNPVLRKHOREITRTTLOSDBEIDYDDTISVEMKEDFDIY 1680  
Qy 787 DEBENQSPRFQKTRHYFIAAVERLWDYQMSSSPHVLRNRAQSGSVPOFKKVVQFQFTD 846  
Db 1681 DEBENQSPRFQKTRHYFIAAVERLWDYQMSSSPHVLRNRAQSGSVPOFKKVVQFQFTD 1740  
Qy 847 GSTQPIYRGELNEHLLGLGPYIRAEVEDNIMVTFRNQASRPYSFYSSLSIYSEEDQROGA 906  
Db 1741 GSTQPIYRGELNEHLLGLGPYIRAEVEDNIMVTFRNQASRPYSFYSSLSIYSEEDQROGA 1800  
Qy 907 EPRKNFVKNETKTYFWKVOHMAPTKDBFCCKAWAYFSDVLEKOVHSLIGPLLCHT 966  
Db 1801 EPRKNFVKNETKTYFWKVOHMAPTKDBFCCKAWAYFSDVLEKOVHSLIGPLLCHT 1860  
Qy 967 NTLNPAHGRQVTVQEFALFETIFDETKSWYFTENMERNCRAPCNIOMEDTFFKENYRPHA 1026  
Db 1861 NTLNPAHGRQVTVQEFALFETIFDETKSWYFTENMERNCRAPCNIOMEDTFFKENYRPHA 1920  
Qy 1027 INGYIMDTLPLVMAQDQIRIRWYLLSMGSENTHSIHFSGHVPTVRKKEBKALYNLYP 1086  
Db 1921 INGYIMDTLPLVMAQDQIRIRWYLLSMGSENTHSIHFSGHVPTVRKKEBKALYNLYP 1980  
Qy 1087 GVPEVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCOTPLGMAHSHIRDFQITAS 1146  
Db 1981 GVPEVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCOTPLGMAHSHIRDFQITAS 2040  
Qy 1147 GOYQWAPKLARLHYSGSINAWSTKEPFSWKIVDLPAMIHIGIKTOGARQKFTSSLYISO 1206  
Db 2041 GOYQWAPKLARLHYSGSINAWSTKEPFSWKIVDLPAMIHIGIKTOGARQKFTSSLYISO 2100  
Qy 1207 FIIMYSLDGKKWQTYRGNSGTGLMVFFGNVDSGIXHNIENPPIIARIYIRLHPHYSIRS 1266  
Db 2101 FIIMYSLDGKKWQTYRGNSGTGLMVFFGNVDSGIXHNIENPPIIARIYIRLHPHYSIRS 2160  
Qy 1267 TLRWELMGCDLNSCMLPGLMESKAISDAQITASSYFTNMFPATWSPSKARHLQGRNAWR 1326  
Db 2161 TLRWELMGCDLNSCMLPGLMESKAISDAQITASSYFTNMFPATWSPSKARHLQGRNAWR 2220  
Qy 1327 POVNNPKMLQVDFQKTMKVTVTQGVKSLTSMYVKEFLISSODGHQWTLFFQNGKV 1386  
Db 2221 POVNNPKMLQVDFQKTMKVTVTQGVKSLTSMYVKEFLISSODGHQWTLFFQNGKV 2280  
Qy 1387 KVFQGNQDSTPVPVNSLDPLLRILYRIHPQSWVHQAIALRMEVLGCEAODLY 1438  
Db 2281 KVFQGNQDSTPVPVNSLDPLLRILYRIHPQSWVHQAIALRMEVLGCEAODLY 2332

RESULT 7

US-10-187-319-2  
; Sequence 2, Application US/10187319  
; Publication No. US20030068785A1  
; GENERAL INFORMATION:  
; APPLICANT: Lollar, John S.  
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII  
; NUMBER OF SEQUENCES: 40  
; CORRESPONDENCE ADDRESS:  
; ADDRESSES: Greenlee, Winner and Sullivan, P.C.  
; STREET: 5370 Manhattan Circle Suite 201  
; CITY: Boulder  
; STATE: Colorado  
; COUNTRY: USA  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10187,319  
; FILING DATE: 27-Aug-2002  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 09/523,656

FILED DATE: 2000-03-10  
APPLICATION NUMBER: US 09/037,601  
FILING DATE: 1998-03-10  
APPLICATION NUMBER: WO PCT/US97/11155  
FILING DATE: 1997-06-26  
APPLICATION NUMBER: US 08/670,707  
FILING DATE: 1996-06-26  
ATTORNEY/AGENT INFORMATION:  
NAME: Greenlee, Lorraine L.  
REGISTRATION NUMBER: 27,894  
REFERENCE/DOCKET NUMBER: 75-95K  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 303/499-8080  
TELEFAX: 303/499-8089  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2332 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: <Unknown>  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
TISSUE TYPE: Liver  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-10-187-319-2

Query Match 94.1%; Score 7234; DB 15; Length 2332;  
Best Local Similarity 61.7%; Pred. No. 0;  
Matches 1438; Conservative 0; Mismatches 0; Indels 894; Gaps 1;  
QY 1 ATRRYLGAVELSWDMSQDLGELPVDAREPPVPKSPFPNTSVVYKKTILFVEFTVHLFN 60  
DB 1 ATRRYLGAVELSWDMSQDLGELPVDAREPPVPKSPFPNTSVVYKKTILFVEFTVHLFN 60  
QY 61 IAKPRPPMGLLGTQAEVYDVITLKNMASHPVSLHAGVSYWKASEGAEYDDQTSQ 120  
DB 61 IAKPRPPMGLLGTQAEVYDVITLKNMASHPVSLHAGVSYWKASEGAEYDDQTSQ 120  
QY 121 REKEDKVPFGSGHTYVQVLKENGPMASDPLCLTYSLSHVDLVKDLNSGLIGALLVCR 180  
DB 121 REKEDKVPFGSGHTYVQVLKENGPMASDPLCLTYSLSHVDLVKDLNSGLIGALLVCR 180  
QY 181 EGSIAKEKTQTLHKFILLFAVDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNR 240  
DB 181 EGSIAKEKTQTLHKFILLFAVDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNR 240  
QY 241 SLPLGICHRKSVVHVIGMTTPEVHSIFLEGHTFLVRNHRQASLEISPIITFLTAQTL 300  
DB 241 SLPLGICHRKSVVHVIGMTTPEVHSIFLEGHTFLVRNHRQASLEISPIITFLTAQTL 300  
QY 301 MDLQQLFLFCHISSHQHDGMEAYVKVDSCPEEPQLRMKNNEAEYDDDLTDSMOWVRF 360  
DB 301 MDLQQLFLFCHISSHQHDGMEAYVKVDSCPEEPQLRMKNNEAEYDDDLTDSMOWVRF 360  
QY 361 DDNSPSFIQIRSVAKGPKTWHYIAAEEDWDYAPLVLPADDRSVKSYQYLANGPORIG 420  
DB 361 DDNSPSFIQIRSVAKGPKTWHYIAAEEDWDYAPLVLPADDRSVKSYQYLANGPORIG 420  
QY 421 RYKVKRFMAVDTDETKTREAIOHESGILGPLYGEVGDITLIIIFKNQASRPYNIYPHGI 480  
DB 421 RYKVKRFMAVDTDETKTREAIOHESGILGPLYGEVGDITLIIIFKNQASRPYNIYPHGI 480  
QY 481 TDVRLPYSRRLPKGVKHLKOPFILLPGEIFKYKWTVTVEDGPKSDPRCLTRYSSFYVME 540  
DB 481 TDVRLPYSRRLPKGVKHLKOPFILLPGEIFKYKWTVTVEDGPKSDPRCLTRYSSFYVME 540  
QY 541 RDLASGLIGPLLIICYKESVDORGNOIMSDKKNVILPSVFDNRNRYWTENIQRELNPAG 600  
DB 541 RDLASGLIGPLLIICYKESVDORGNOIMSDKKNVILPSVFDNRNRYWTENIQRELNPAG 600

QY 601 VQLEDPEFOASNIHMSINGYVFDLSQLSVCLHEVAYWYILSGAQTDFLSVFFSGYTPKH 660  
DB 601 VQLEDPEFOASNIHMSINGYVFDLSQLSVCLHEVAYWYILSGAQTDFLSVFFSGYTPKH 660  
QY 661 KMVYEDTTLTLPFGSETVFMSENPGLWILGCHNSDFRNRGWTALLKYSSCDKNTGDYYE 720  
DB 661 KMVYEDTTLTLPFGSETVFMSENPGLWILGCHNSDFRNRGWTALLKYSSCDKNTGDYYE 720  
QY 721 DSYEDISAYLLSKNNAIEPRSF----- 742  
DB 721 DSYEDISAYLLSKNNAIEPRSFQNSRHPSTROKFNATTIPENDIETKDPWFAHRTMP 780  
QY 743 ----- 742  
DB 781 KIQNVSSDLLMLLRQSPTPHGLSLSLDLQEAKEYTFSDDPSPGAIDSNNSISEMTHFRPQ 840  
QY 743 ----- 742  
DB 841 LHHSGDMVTPESGLQLRLNEKLGTAAATELKKLDFKVSSSTNNLITIPSDNLAAGTDN 900  
QY 743 ----- 742  
DB 901 TSSLGPSPMPVHYDSQLDTTLFGKSSPLTESGGPLSLSEENNDKLLSGLMNSQESSW 960  
QY 743 ----- 742  
DB 961 GKNVSTESGRLFKGRAGHPALLTKDNALFKVSI SLLKTNKTSNNSATNRKTHIDGPSL 1020  
QY 743 ----- 742  
DB 1021 LIENSFSVWQNIESTDFPKVTPLIHDRMLMDKNATALLNHNMSKNTSSKNMEMVQOK 1080  
QY 743 ----- 742  
DB 1081 KEGPIPPDAQNDMSFFKMLFLPESARWITQTHGKNSLNSGQSPSPKQLVSLGPEKSVBG 1140  
QY 743 ----- 742  
DB 1141 QNFLSEKNKVVGKGEFTKDVGLKEMVFPSPSRNLFITNLDLHNHNTHNQEKIOBEIEK 1200  
QY 743 ----- 742  
DB 1201 KETLIQENVVLPIQIHTVTGKFMKMLFLLSTRQNVGSEYEGAYAPVLQDFRSLNDSTNR 1260  
QY 743 ----- 742  
DB 1261 TKKHTAHSKGBEENLEGLGNQTKQIVEKYACTTRI SPNTSQNFVTOQSKRALQFRL 1320  
QY 743 ----- 742  
DB 1321 PLEETELEKRIIIVDDTSTQSKNMKHLTPSTLTQIDYNEKKGAIQTSPLSDCLTRSHSI 1380  
QY 743 ----- 742  
DB 1381 PQANRSLPIAKVSSPSPRIPIYLITRVLFDQNSSHLPAASRYKDKSGVQESSHFLOGAKK 1440  
QY 743 ----- 742  
DB 1441 NNLSLAILTLEMTGDQREVGLSTSATNSVYKKVENTVLPKPDLPKTSKGVKELLPKVHI 1500  
QY 743 ----- 742  
DB 1501 YQKDLFPFTETNSGPHGLDLVEGSLAQTEGAIKMNEANRPKGVPLRVATESSAKTPSK 1560  
QY 743 ----- 742  
DB 1561 LLDPLAWDNHYGTQIPKEBWKSOEKSPEKTAFFKKDTIISLNACENHAI AAINEQNKP 1620  
QY 743 ----- 786  
DB 1621 EIEVTWAKQORTRELCSQNPPLKRRHQREITRTTLQSDQDEIIDYDDTISVEMKEDFDIY 1680



QY 787 DEBENQSPRFOKKTIRHYFTAAVERLWDYQMSSSPHVLRNRAQSGSVPOKQKVVFOEFTD 846  
Db 1681 DEBENQSPRFOKKTIRHYFTAAVERLWDYQMSSSPHVLRNRAQSGSVPOKQKVVFOEFTD 1740  
QY 847 GSTOPLYRGELNEHLGLGPYIRAEVEDNIMWTFRNQASRPYSFYSSLLSYSEEDORQGA 906  
Db 1741 GSTOPLYRGELNEHLGLGPYIRAEVEDNIMWTFRNQASRPYSFYSSLLSYSEEDORQGA 1800  
QY 907 EPRKNFVKNETKTYFMKVOHMAPTKDEEDCKAWAYFSDVDLEKOVHSGLIGPLLVCHT 966  
Db 1801 EPRKNFVKNETKTYFMKVOHMAPTKDEEDCKAWAYFSDVDLEKOVHSGLIGPLLVCHT 1860  
QY 967 NTLNPAHGRQVTVQEFALFTTIDEKTSWYFTENMERNCRAPCNIQMEDPTFKENTRPHA 1026  
Db 1861 NTLNPAHGRQVTVQEFALFTTIDEKTSWYFTENMERNCRAPCNIQMEDPTFKENTRPHA 1920  
QY 1027 INGVIYDMLPGLVMAQDORIRWYLLSMGNSNENIHSIHFSGHVFTVRKEEYKVALYNLYP 1086  
Db 1921 INGVIYDMLPGLVMAQDORIRWYLLSMGNSNENIHSIHFSGHVFTVRKEEYKVALYNLYP 1980  
QY 1087 GVFTVEMLPSKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPGLGMASGHIRDFQITAS 1146  
Db 1981 GVFTVEMLPSKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPGLGMASGHIRDFQITAS 2040  
QY 1147 GOYGOWAPKARLHYSGSINAWSTKEPPSWIKVDLLAPMIHIGIKTOGARQKPSLSYISQ 1206  
Db 2041 GOYGOWAPKARLHYSGSINAWSTKEPPSWIKVDLLAPMIHIGIKTOGARQKPSLSYISQ 2100  
QY 1207 FILMYSLDGKKWOTYRGNSTGLMVFPNGVDSGIKHNIENPPIIARYIRLHPTHYSIRS 1266  
Db 2101 FILMYSLDGKKWOTYRGNSTGLMVFPNGVDSGIKHNIENPPIIARYIRLHPTHYSIRS 2160  
QY 1267 TLRMELMGCDLNSCSMPLGMSKAISSAQITASSYFTNMFATWSPSKARLHLQGRSNAMR 1326  
Db 2161 TLRMELMGCDLNSCSMPLGMSKAISSAQITASSYFTNMFATWSPSKARLHLQGRSNAMR 2220  
QY 1327 POWNPKEWLQVDFQKTMKVTVTQGVKSLLTSVMYKEPLISSQDGHQWTLFPQNGKV 1386  
Db 2221 POWNPKEWLQVDFQKTMKVTVTQGVKSLLTSVMYKEPLISSQDGHQWTLFPQNGKV 2280  
QY 1387 KVFQGNQDSPTPVVNSLDPLLRYLRIHPQSVWQIALRMEVLGCEAODLY 1438  
Db 2281 KVFQGNQDSPTPVVNSLDPLLRYLRIHPQSVWQIALRMEVLGCEAODLY 2332

## RESULT 8

US-10-133-907-4  
; Sequence 4, Application US/10133907  
; Publication No. US20030195223A1  
; GENERAL INFORMATION:  
; APPLICANT: Chien, Kenneth R  
; APPLICANT: Hoshijima, Masahiko  
; TITLE OF INVENTION: Method to treat hemophilia by hepatic gene transfer of Factor VII  
; TITLE OF INVENTION: with vesicle vector  
; FILE REFERENCE: 6627-PAL170  
; CURRENT APPLICATION NUMBER: US/10/133,907  
; PRIOR FILING DATE: 2002-04-25  
; PRIOR APPLICATION NUMBER: 60/286,314  
; PRIOR FILING DATE: 2001-04-25  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 4  
; LENGTH: 2351  
; TYPE: PRN  
; ORGANISM: Homo sapiens  
US-10-133-907-4

Query Match 94.0%; Score 7227; DB 12; Length 2351;  
Best Local Similarity 61.6%; Pred. No. 0;  
Matches 1437; Conservative 0; Mismatches 1; Indels 894; Gaps 1;  
QY 1 ATRRYLGAVELSWDYNQSDLGELPVDARPPRPVKSPFNTSVVYKTKLTFVEFTVHLEN 60  
|||||

Db 20 ATRRYLGAVELSWDYNQSDLGELPVDARPPRPVKSPFNTSVVYKTKLTFVEFTVHLEN 79  
QY 61 IAKPRPFWMLGLLPTTQAEVYDVTVTITLKNMASHPVSLHAGVSYWKASEGAYDDQTSQ 120  
Db 80 IAKPRPFWMLGLLPTTQAEVYDVTVTITLKNMASHPVSLHAGVSYWKASEGAYDDQTSQ 139  
QY 121 REKEDDKVPFGSGHTTVMQVLYKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 180  
Db 140 REKEDDKVPFGSGHTTVMQVLYKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 199  
QY 181 EGS LAKEKTQTLHKFTLLFAVFDGKSWHSETKNSLMODRDAASAPAKPMHTVNGVNR 240  
Db 200 EGS LAKEKTQTLHKFTLLFAVFDGKSWHSETKNSLMODRDAASAPAKPMHTVNGVNR 259  
QY 241 SLFGLIGCHRKSYVMHVIQMGTTTPEVHSIFLEGTFLVRNHRQASLEISPIITFLTAQTLL 300  
Db 260 SLFGLIGCHRKSYVMHVIQMGTTTPEVHSIFLEGTFLVRNHRQASLEISPIITFLTAQTLL 319  
QY 301 MDLGOPLLCHISSHODGMEAYVKVDSCEPQPOLMKNEEAEDYDDDLTSEMDEVRF 360  
Db 320 MDLGOPLLCHISSHODGMEAYVKVDSCEPQPOLMKNEEAEDYDDDLTSEMDEVRF 379  
QY 361 DDNSPSFIQIRSVAKKHPTWVHYIAAEEEDWDYAPLVADDDRYSKYQLNNGPQIRG 420  
Db 380 DDNSPSFIQIRSVAKKHPTWVHYIAAEEEDWDYAPLVADDDRYSKYQLNNGPQIRG 439  
QY 421 RYKVKYRFMAYTDETFKTRERAIQHESGILGPLLYGEVGTLLIFKNQASRPNIYPHGI 480  
Db 440 RYKVKYRFMAYTDETFKTRERAIQHESGILGPLLYGEVGTLLIFKNQASRPNIYPHGI 499  
QY 481 TVRPLYSRRLPGVKHLKDFPILPGEIFKYKWTVTVEGPTKSDPRCLTRYSSPVNME 540  
Db 500 TVRPLYSRRLPGVKHLKDFPILPGEIFKYKWTVTVEGPTKSDPRCLTRYSSPVNME 559  
QY 541 RDLASGLIGPLLI CYKESVDQRCNOIMSDKNVILFESVDENESWYLTENIQBFLENPAG 600  
Db 560 RDLASGLIGPLLI CYKESVDQRCNOIMSDKNVILFESVDENESWYLTENIQBFLENPAG 619  
QY 601 VOLEDEPFOASINMHSINGVYFDSQLSVCLHEVAYWYILSIGAQDTFLSVFSSGTFPKH 660  
Db 620 VOLEDEPFOASINMHSINGVYFDSQLSVCLHEVAYWYILSIGAQDTFLSVFSSGTFPKH 679  
QY 661 KMVYEDTLTLFPFSGETVFMENPGLWILGCHNSDFRNRGMTALLKVSSCDKNTGDYFE 720  
Db 680 KMVYEDTLTLFPFSGETVFMENPGLWILGCHNSDFRNRGMTALLKVSSCDKNTGDYFE 739  
QY 721 DSYEDISAYLLSKNAIEPRSF ----- 742  
Db 740 DSYEDISAYLLSKNAIEPRSFQNSRHPSTRQKQFNATTIPENDIEKTDPNFAHRTMP 799  
QY 743 ----- 742  
Db 800 KIQNVSSDILLMLLRSPTPHGLSLSDLOEAKVETSDPSPCAIDNNSLSNMTFRPQ 859  
QY 743 ----- 742  
Db 860 LHHSGDMWTPPSGLQLRLNEKLGTTAATLKLDPKVSSTNNLSTIPSDMLAAGTDN 919  
QY 743 ----- 742  
Db 920 TSSLGPPSPMVDYHVSOLDTTLFGKSSPLTESGGPLSLEENNDKSLFSLGMLNSQESSW 979  
QY 743 ----- 742  
Db 980 GKNVSTESGRLFKGKRAHGPALLTKDNALFKVSI LKTKNTKSNNSATNRKTHIDGPSL 1039  
QY 743 ----- 742  
Db 1040 LIENSPVQNLILESDTEFKVYTFPLIHDRMLDKNATLRLNHSNKTSSKNWEMVQK 1099  
QY 743 ----- 742  
Db 1100 KEGPIPPDAQNDPMSFFKMLFLPESARWIQRTGKNSLNSGQSPKQLVSLGPEKSV 1159

QY 743 ----- 742  
Db 1160 QNFLSEKXNVVGKEFTKQVLGKEMVPPSSRNLFNLNHLNHNTHNQEKIOEBIEK 1219  
QY 743 ----- 742  
Db 1220 KETLIQENVLPQIHVTGTGKXFNKULFLLSTRQNVESYDGAVALQDFRSLNDSTNR 1279  
QY 743 ----- 742  
Db 1280 TKHTAHFSKKGBENLEGLGNQTKQIVEKYACTTRISPNTSQNFVTVQSRKALKQPL 1339  
QY 743 ----- 742  
Db 1340 PLEETELEKRIIVDDTSTQWKNMKHLTPSTLTQIDYNEKEKGAITQSPDLCULTRSHI 1399  
QY 743 ----- 742  
Db 1400 PQANRSLPIAKVSFPSPIRPIYLTRVLFDQNSHSLPAASYRKDQSGVQBSHFLQAKK 1459  
QY 743 ----- 742  
Db 1460 NNLSAILTLEMTGDQREVSGISGATNSVTYKVENTVLPKPLPKTSGKVLLPKVHI 1519  
QY 743 ----- 742  
Db 1520 YQKDLFPPTETNSGPHLDLVEGSLLOQTEGAIKWNEANRPGKVPFLRVATESSAKTPSK 1579  
QY 743 ----- 742  
Db 1580 LLDPLAWNHYGTQIPKEBWSQEKSPKTAFFKXDTTILSNACSNHAIAINBQGNPK 1639  
QY 743 ----- 742  
Db 1640 EIEVTWAKQGRTERLCSQNPVLKRHOREITRTTQSDQEEIDYDDTISVENKKEDFDIY 1699  
QY 787 DEDEQSPRSFQKTRHYFIAVERLDYGMSSPHVLNRAQSGVPQFKVWQEFID 846  
Db 1700 DEDEQSPRSFQKTRHYFIAVERLDYGMSSPHVLNRAQSGVPQFKVWQEFID 1759  
QY 847 GSFTQPLYRGELNEHLGLGPIYRAEVEDNIWTFRNQASRPYSFYSSLISSIEEDQROGA 906  
Db 1760 GSFTQPLYRGELNEHLGLGPIYRAEVEDNIWTFRNQASRPYSFYSSLISSIEEDQROGA 1819  
QY 907 EPRKPFVKNETKTYFWKQHNMAPTKDBFCCKAWAYFSDVLEKDVHSGLIGPLLVCIT 966  
Db 1820 EPRKPFVKNETKTYFWKQHNMAPTKDBFCCKAWAYFSDVLEKDVHSGLIGPLLVCIT 1879  
QY 967 NTLNPAHGRQVTVQEEALPFTIETDKSWYFTENMERNCRAPCNQIOMEDPTFKENYRFA 1026  
Db 1880 NTLNPAHGRQVTVQEEALPFTIETDKSWYFTENMERNCRAPCNQIOMEDPTFKENYRFA 1939  
QY 1027 INGIYMDTLPGLVMAQDQRIWYLLSGNSNENIHSIHFGSHVFTVRKBEYKVALYNLYP 1086  
Db 1940 INGIYMDTLPGLVMAQDQRIWYLLSGNSNENIHSIHFGSHVFTVRKBEYKVALYNLYP 1999  
QY 1087 GVFTVEMLPKAGIWRVECLIGEHLHAGMSTFLVYGNKQCPPLGMAASHIRDFQITAS 1146  
Db 2000 GVFTVEMLPKAGIWRVECLIGEHLHAGMSTFLVYGNKQCPPLGMAASHIRDFQITAS 2059  
QY 1147 GOYGOWAPKLARLHYSGINAWSTKEPFSWKVDLLAPMIIHGIKTCQARQKFSYLSIQ 1206  
Db 2060 GOYGOWAPKLARLHYSGINAWSTKEPFSWKVDLLAPMIIHGIKTCQARQKFSYLSIQ 2119  
QY 1207 FIWYSDGKKWQYRNGSTGTLVFFGNVDSGKHNINFPPIIARYIRLHPHYSIRS 1266  
Db 2120 FIWYSDGKKWQYRNGSTGTLVFFGNVDSGKHNINFPPIIARYIRLHPHYSIRS 2179  
QY 1267 TLRMELMGCDLNSCSMPLGMESKAISDAQITASSYFTNMFPATWSPSKARLHLOGRSNAR 1326  
Db 2180 TLRMELMGCDLNSCSMPLGMESKAISDAQITASSYFTNMFPATWSPSKARLHLOGRSNAR 2239

QY 1327 PQVNNPKEMLOVDFQKTMKVGTGTVQVKSLLTSMYKBEFLISSQDCHQWTLFFQNGKV 1386  
Db 2240 PQVNNPKEMLOVDFQKTMKVGTGTVQVKSLLTSMYKBEFLISSQDCHQWTLFFQNGKV 2299  
QY 1387 KVFQGNDSFTPVVNSIDPPLLTRYLRIHPQSWHQAIALRMEVLGCEAODLY 1438  
Db 2300 KVFQGNDSFTPVVNSIDPPLLTRYLRIHPQSWHQAIALRMEVLGCEAODLY 2351  
RESULT 9  
US-10-132-829-4  
; Sequence 4, Application US/10132829  
; Publication No. US2003004982A1  
; GENERAL INFORMATION:  
; APPLICANT: Chien, Kenneth R  
; APPLICANT: Hoshijima, Masahiko  
; TITLE OF INVENTION: Method to treat hemophilia by hepatic gene transfer of Factor V  
; TITLE OF INVENTION: with vesicle vector  
; FILE REFERENCE: 6627-PALL70  
; CURRENT APPLICATION NUMBER: US/10/132,829  
; CURRENT FILING DATE: 2002-04-25  
; PRIOR APPLICATION NUMBER: 60/286,314  
; PRIOR FILING DATE: 2001-04-25  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 4  
; LENGTH: 2351  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-132-829-4  
Query Match 94.0%; Score 7227; DB 15; Length 2351;  
Best Local Similarity 61.6%; Pred. No. 0;  
Matches 1437; Conservative 0; Mismatches 1; Indels 894; Gaps 1;  
QY 1 ATRRYYLGAVELSWDYMOSDIEGELPYDARFPVRPKSPFNSTSVYKKTFLVEFTVHLFN 60  
Db 20 ATRRYYLGAVELSWDYMOSDIEGELPYDARFPVRPKSPFNSTSVYKKTFLVEFTVHLFN 79  
QY 61 IAKPRPPMGLLGPTIOAEVDTVVITLKNASHPVSLHVGVSYKASEGAEYDDQTSQ 120  
Db 80 IAKPRPPMGLLGPTIOAEVDTVVITLKNASHPVSLHVGVSYKASEGAEYDDQTSQ 139  
QY 121 REKEDDKVFPQGSHTYVWQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 180  
Db 140 REKEDDKVFPQGSHTYVWQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 199  
QY 181 EGSIAKEKTOHLKILLFAVPEDEKSWHSETKNSLMQDRDAASARAMPKMTVNGYVNR 240  
Db 200 EGSIAKEKTOHLKILLFAVPEDEKSWHSETKNSLMQDRDAASARAMPKMTVNGYVNR 259  
QY 241 SLPLGLIGCHRKSVYWHVIGMGTTPVHSIFLEGHTFLVNRHQASLEISPIITFLTAQTLL 300  
Db 260 SLPLGLIGCHRKSVYWHVIGMGTTPVHSIFLEGHTFLVNRHQASLEISPIITFLTAQTLL 319  
QY 301 MDLQGFLLFCHTSSHQHDMGEAYVVDSCPEPQPLRMKNNEAREYDDDLTDTSEMDVVRP 360  
Db 320 MDLQGFLLFCHTSSHQHDMGEAYVVDSCPEPQPLRMKNNEAREYDDDLTDTSEMDVVRP 379  
QY 361 DDNSPSPFTQIRSAVKAGHPKTHVHTAAEEEDWDYAPLVADPDRSYKSOYLNGPQRTG 420  
Db 380 DDNSPSPFTQIRSAVKAGHPKTHVHTAAEEEDWDYAPLVADPDRSYKSOYLNGPQRTG 439  
QY 421 RYKVKVRFWAYTDTFTKTEALQHSGLIGLPGLLYGEVGDTLIIIFKQASRPYNIYPHGI 480  
Db 440 RYKVKVRFWAYTDTFTKTEALQHSGLIGLPGLLYGEVGDTLIIIFKQASRPYNIYPHGI 499  
QY 481 TVRPLYSRRLPKGVKHLXDPFLLPGEIFKYKWTVTVEDGPKSDPCRTYRYSFVNME 540  
Db 500 TVRPLYSRRLPKGVKHLXDPFLLPGEIFKYKWTVTVEDGPKSDPCRTYRYSFVNME 559  
QY 541 RDLASGLIGLPGLLYCYKESVDQGNQIMSDKNVILFSPVFDENRSWTLTENIORFLPNPAG 600

Db 560 RDLASGLIGLLICYKESVDQRGNQMSDKRNVILSVDFDENRSWLTENIQRLPNPAG 619  
Qy 601 VQLEDFEQASNMHSINGVFDLSQLSVCLHEVAYWYILSIGAQTDPLSVFSGYTFKH 660  
Db 620 VQLEDFEQASNMHSINGVFDLSQLSVCLHEVAYWYILSIGAQTDPLSVFSGYTFKH 679  
Qy 661 KMVEDTLTLPFSGETVFMENPGLWILGCHNSDFRNRGWTALLKVSSCDKNTGDIYE 720  
Db 680 KMVEDTLTLPFSGETVFMENPGLWILGCHNSDFRNRGWTALLKVSSCDKNTGDIYE 739  
Qy 721 DSYEDISAYLLSKNNATEPRSF----- 742  
Db 740 DSYEDISAYLLSKNNATEPRSFQNSRHPSTRQKQFNATTIPENDIEKTDPFARHTMP 799  
Qy 743 ----- 742  
Db 800 KIQNVSSDLLMLLRQSPTEPHGLSLDLQAKYETFDPSGAGIDSNNSISEMTHFRPQ 859  
Qy 743 ----- 742  
Db 860 LHHSGDMVFTPEGLQLRLNEKLGTTAATBLKKLDFKVSSTNNLISITIPSDNLAAGTDN 919  
Qy 743 ----- 742  
Db 920 TSSLGPPSPMPVHYDQDLDTLFGKSSPLTESGGPLSLEENNDKLLSGLMNSQBSW 979  
Qy 743 ----- 742  
Db 980 GKNVSTESGRLFKGRAHPALLTKDNALFKVSI SLLKTNKTSNNSATNRKTHIDGPSL 1039  
Qy 743 ----- 742  
Db 1040 LIENSPSVWONILESTFEKKVTPLIHRLMLMDKNATALRLNHSNKTTSKNMEMVQK 1099  
Qy 743 ----- 742  
Db 1100 KEGPIPPDAQNPDMSFFKMLFLPESARWI ORTHGKNSLNSGQSPKQLVSLGPEKSVEG 1159  
Qy 743 ----- 742  
Db 1160 QNFLSEKNKVVGKEFTKDVGLKEMVPPSRNLFLTWLNLHENNTHNQEKTKQERIEK 1219  
Qy 743 ----- 742  
Db 1220 KETLIOENVVLPQIHVTGTKNFMKNLFLSTRQNVESYGAVAPVLQDRSLNDSTNR 1279  
Qy 743 ----- 742  
Db 1280 TKKHTAHFSKKGBEENLEGLGNQTKQIVEKYACTTRISPNTSQQNFVTRQSRKALKQFRL 1339  
Qy 743 ----- 742  
Db 1340 PLEETELEKRIIVDDTSTQSKNMKHLTPSLTQIDYNEKEGALTQSPSLDCLTRSHSI 1399  
Qy 743 ----- 742  
Db 1400 PQANRSLPIAKVSSFSIRPIYLRVLFDNSSHLPAAVYRKXDSGVQBSHFLQAKK 1459  
Qy 743 ----- 742  
Db 1460 NNLSLAILTLEMTGDREVGSLGTSATNSVYTKKVENTVLPKPLPKTSKGVKVELLPKVI 1519  
Qy 743 ----- 742  
Db 1520 YQKDLFPPTETSGSPGHLDLVEGSLLOQTEGAIKWNEARPKVPFLRVATESAKTPSK 1579  
Qy 743 ----- 742  
Db 1580 LLDPLAWDNHYGTQIPKEWKSQEKSPKTAPEKKKDTILSLNACESNHAIAINEGQWKP 1639  
Qy 743 -----SONPVLKXHQREITRTTLOSQOEIDYDDTISVEMKKEDFDIY 786  
Db 1640 EIEVTWAKQGRTERLCSQNPVLKXHQREITRTTLOSQOEIDYDDTISVEMKKEDFDIY 1699

Qy 787 DEENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNRAGSGVPPQKKVVFQBFETD 846  
Db 1700 DEENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNRAGSGVPPQKKVVFQBFETD 1759  
Qy 847 GSTQPLYRGELNEHLLGLPGYIRAEVEDNIMVTERNOASRPYSFYSSLSIYEDORQGA 906  
Db 1760 GSTQPLYRGELNEHLLGLPGYIRAEVEDNIMVTERNOASRPYSFYSSLSIYEDORQGA 1819  
Qy 907 EPRKNFVKNETKTYFWKVQHMAPTKDEFCKAMAYFSDVLEKDVHSGSLIGPLLVCHT 966  
Db 1820 EPRKNFVKNETKTYFWKVQHMAPTKDEFCKAMAYFSDVLEKDVHSGSLIGPLLVCHT 1879  
Qy 967 NTLNPAHGRQVTVQEFALPFTTIDETKSWYFTFMERNCRAPCNIOHEDPTFKENTRPHA 1026  
Db 1880 NTLNPAHGRQVTVQEFALPFTTIDETKSWYFTFMERNCRAPCNIOHEDPTFKENTRPHA 1939  
Qy 1027 INGYIMDTLPLGLVMAQDQIRWYLLSGNSNENHSHFSGHVFTRKCKEYKVALYNLYP 1086  
Db 1940 INGYIMDTLPLGLVMAQDQIRWYLLSGNSNENHSHFSGHVFTRKCKEYKVALYNLYP 1999  
Qy 1087 GFVETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPPLGWSGHIRDFQITAS 1146  
Db 2000 GFVETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPPLGWSGHIRDFQITAS 2059  
Qy 1147 GOYGWAPKARLHYSGSINAWSTKBPFSWIKVDLLAPMIHIGIKTOGARQKFSLSYSQ 1206  
Db 2060 GOYGWAPKARLHYSGSINAWSTKBPFSWIKVDLLAPMIHIGIKTOGARQKFSLSYSQ 2119  
Qy 1207 FIIMYSIDGKKWOTYRGNSGTGLWVFGNVDSGIXHNIENPPIIARYIRLHPHYSIRS 1266  
Db 2120 FIIMYSIDGKKWOTYRGNSGTGLWVFGNVDSGIXHNIENPPIIARYIRLHPHYSIRS 2179  
Qy 1267 TLRMELMGCDLNSCMLPLGWSKAI SDAQITASSYFTNMPATWSPSKARLHLOQSNWR 1326  
Db 2180 TLRMELMGCDLNSCMLPLGWSKAI SDAQITASSYFTNMPATWSPSKARLHLOQSNWR 2239  
Qy 1327 PQVNNPKWLOVDPOKTMKVTVGTQGVKSLLTSMYVKEFLISSQDGHQWTLFFQNGKV 1386  
Db 2240 PQVNNPKWLOVDPOKTMKVTVGTQGVKSLLTSMYVKEFLISSQDGHQWTLFFQNGKV 2299  
Qy 1387 KVFQGNQDSPTPVVNSLDPELLTRYLRIRHPQSVWVHQTALRMEVLGCEAQDLY 1438  
Db 2300 KVFQGNQDSPTPVVNSLDPELLTRYLRIRHPQSVWVHQTALRMEVLGCEAQDLY 2351  
RESULT 10  
US-10-172-712-27  
; Sequence 27, Application US/10172712  
; Publication No. US20030125232A1  
; GENERAL INFORMATION:  
; APPLICANT: GRIFFIN, JOHN H.  
; APPLICANT: GALE, ANDREW J.  
; APPLICANT: GETZOFF, ELIZABETH D.  
; APPLICANT: FELLEQUER, JEAN-LUC  
; TITLE OF INVENTION: STABILIZED PROTEINS WITH ENGINEERED DISULFIDE BONDS  
; FILE REFERENCE: 4198-4001US1  
; CURRENT APPLICATION NUMBER: US/10172,712  
; CURRENT FILING DATE: 2002-09-30  
; PRIOR APPLICATION NUMBER: 60/238,578  
; PRIOR FILING DATE: 2001-06-14  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 27  
; LENGTH: 2351  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-172-712-27

Query Match 94.0%; Score 7227; DB 15; Length 2351;  
Best Local Similarity 61.6%; Pred. No. 0;  
Matches 1437; Conservative 0; Mismatches 1; Indels 894; Gaps 1;

QY	1	ATRRYVLGAVELSDYMSDGLGELPVDARPPRPVPSPPFNTSVVYKKTFLVETVHLEN	60
Db	20	ATRRYVLGAVELSDYMSDGLGELPVDARPPRPVPSPPFNTSVVYKKTFLVETVHLEN	79
QY	61	IAKPRPPMGLLGPTIQAEVYDVTVITLKNMASHPVSLHVGVSYWKASGAEYDDOTSQ	120
Db	80	IAKPRPPMGLLGPTIQAEVYDVTVITLKNMASHPVSLHVGVSYWKASGAEYDDOTSQ	139
QY	121	REKEDKVPFGSHYVQVVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR	180
Db	140	REKEDKVPFGSHYVQVVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR	199
QY	181	EGSLAKEKTOTLHKILLPAVDEGKSWHSETKNSLMODRDAASARAMPKMHYVNGYNR	240
Db	200	EGSLAKEKTOTLHKILLPAVDEGKSWHSETKNSLMODRDAASARAMPKMHYVNGYNR	259
QY	241	SLPGLIGCHRSKVYVHVTGMGTPPEVHSIFLEGHTFLVRNHRQASLEISPIITFLTAQTL	300
Db	260	SLPGLIGCHRSKVYVHVTGMGTPPEVHSIFLEGHTFLVRNHRQASLEISPIITFLTAQTL	319
QY	301	MDLGQFLLPCHTSSHOHDGMEAYVVDSCPEPQOLRMKNNEAEYDDDLTDSMDVVRP	360
Db	320	MDLGQFLLPCHTSSHOHDGMEAYVVDSCPEPQOLRMKNNEAEYDDDLTDSMDVVRP	379
QY	361	DDNSPSFTQIRSVAKKPKTWVHYIAAEEEDWDYAPLVLAAPDRSRYKSOYLNNQPORIG	420
Db	380	DDNSPSFTQIRSVAKKPKTWVHYIAAEEEDWDYAPLVLAAPDRSRYKSOYLNNQPORIG	439
QY	421	RKYKVRPMAYTDTFTKTRAIQHESGILGPLYGEVGTLLIIFPNQASRYNIYPHGI	480
Db	440	RKYKVRPMAYTDTFTKTRAIQHESGILGPLYGEVGTLLIIFPNQASRYNIYPHGI	499
QY	481	TDVRPLYRRLPKGVKHLKDFPILPGEIIFYKWTVTVEGPTKSDPRCLTRYSSVFNME	540
Db	500	TDVRPLYRRLPKGVKHLKDFPILPGEIIFYKWTVTVEGPTKSDPRCLTRYSSVFNME	559
QY	541	RLASGLIGPLTICYKESVDQRGNQJMSDKRNVILFSVDENRSWYLTENIQRFLENPAG	600
Db	560	RLASGLIGPLTICYKESVDQRGNQJMSDKRNVILFSVDENRSWYLTENIQRFLENPAG	619
QY	601	VOLEDEFOASNMHISINGVVFDSQLSVCLHEVAYWYILSIGAQTDFLSVFPSTFKH	660
Db	620	VOLEDEFOASNMHISINGVVFDSQLSVCLHEVAYWYILSIGAQTDFLSVFPSTFKH	679
QY	661	KMYEDTLTLPFPGSETVFMENPGLWILGCHNSDFRNGMTALLKVSJCDKNTGDYIE	720
Db	680	KMYEDTLTLPFPGSETVFMENPGLWILGCHNSDFRNGMTALLKVSJCDKNTGDYIE	739
QY	721	DSYEDISAYLLSKNNAIEPRSF-----	742
Db	740	DSYEDISAYLLSKNNAIEPRSFQNSRHPSTROKQFNATIPENDIEKTDPPFAHRTMP	799
QY	743	-----	742
Db	800	KIQNVSSDILLMLLRQSPHGLSLDLQEAKEYTFSDPSPGALDNNLSLSEMTFRPQ	859
QY	743	-----	742
Db	860	LHSGDMVFTPEGLQRLINEXLGTTAATELKKLDFKVSSTSNLITIPSDNLAAGTDN	919
QY	743	-----	742
Db	920	TSSLGPPMPVHYDSQDITTLFGKSSPLTESGGPLSBEENNDKSLLESGLMNSQESSW	979
QY	743	-----	742
Db	980	GKNVSTESGELFKGKRAHGPAALTNDNALFKVSIILLKTNKTSNNSATNRKTHIDGPSL	1039
QY	743	-----	742
Db	1040	LIENSPSVWQNIESTDETFKKVVTPLIHDRMLMDKNATALRLNHNKNTSSKNMVMQOK	1099
QY	743	-----	742
Db	1100	KEGIPPPDAQNPOMSPFFKMLFLPESARWIORTHGKNSLNSGQSPKQVLSLGEKSVBG	1159
QY	743	-----	742
Db	1160	QNFLEKKNVVVGKEFTKDVGLKEMVFPSSRNLFLTNLDNLHNNTHNQEKIIEBIEK	1219
QY	743	-----	742
Db	1220	KETLIQENVVLPQIHVTGTGKNFMKNLFLLLSTRQNVGSDYGAVPLQDFRSLDSTNR	1279
QY	743	-----	742
Db	1280	TKKHTAHSFKGEBENLEGLNQTKQVEKYVACTTRISPNTSQONFVQSRKRALKQFRL	1339
QY	743	-----	742
Db	1340	PLEBTELEKRIIVDDTSTQMSKNMKHLTPSLTQIDYNEKEKGALITQSPSLDCLTRSHI	1399
QY	743	-----	742
Db	1400	POANRSLPIAKVSFSPSIRPIYLRVLFDQNSSHLPAASRYKKDSGVQBSHFLQGAOK	1459
QY	743	-----	742
Db	1460	NNLSAILTLBMTGDQREVSGLSATSNTSVYKKVENTVLPKPDLPKTSGKVELLPKVI	1519
QY	743	-----	742
Db	1520	YQKDLFPPTETSNGPSGLDLVEGSLLOQTEGAIKWNEANRPGKVPFLRVATESSAKTPSK	1579
QY	743	-----	742
Db	1580	LILDPLAWNHYGTQIPKEBWSQEKSPKTAFFKKDDTILSLNACSNHAIAINSGONKP	1639
QY	743	-----SONPPVLKEHOREITRTTLQSQOEBTDYDDTISVEMKKEDFDIY	786
Db	1640	EIEVTWAKQGRTERLCSQNPVLRKHOREITRTTLQSQOEBTDYDDTISVEMKKEDFDIY	1699
QY	787	DDENQSPRSFOKTRHYFIAAVERLDWYGMSSSPHVLNRAQSGSVPOFKVVFQEFTH	846
Db	1700	DDENQSPRSFOKTRHYFIAAVERLDWYGMSSSPHVLNRAQSGSVPOFKVVFQEFTH	1759
QY	847	GSFTQPLYRGELENEHLLGPIYIRAEVDNIIMVTRNQASRPYSYFSSLSIYEEBQOQA	906
Db	1760	GSFTQPLYRGELENEHLLGPIYIRAEVDNIIMVTRNQASRPYSYFSSLSIYEEBQOQA	1819
QY	907	EPKRFVKNETKTYFWKQHEHMAPTKDFDCKAWAYPSDVKDQVHSGLIGPLIVCHT	966
Db	1820	EPKRFVKNETKTYFWKQHEHMAPTKDFDCKAWAYPSDVKDQVHSGLIGPLIVCHT	1879
QY	967	NTLNPAGRQVTVQEPALPFTTIFDETKSWYFTENMERNCRAPCNIQMEDPTPKENYRFA	1026
Db	1880	NTLNPAGRQVTVQEPALPFTTIFDETKSWYFTENMERNCRAPCNIQMEDPTPKENYRFA	1939
QY	1027	INGYIMDTLPLGLVMAQDQIRWYLLSGNENIHSIHFSGHVTTPKKEEYKXALYNLYP	1086
Db	1940	INGYIMDTLPLGLVMAQDQIRWYLLSGNENIHSIHFSGHVTTPKKEEYKXALYNLYP	1999
QY	1087	GUPETVEMLPKAGIWRVCLIGEHLHAGMSTLFLVYNNKOTPLGASCHIDFOITAS	1146
Db	2000	GUPETVEMLPKAGIWRVCLIGEHLHAGMSTLFLVYNNKOTPLGASCHIDFOITAS	2059
QY	1147	GOYGOWAPKLARLHYSGSINAMSTKEPFSWIKVYDLAPMIITHIGIKTQAGRKQFSSYISQ	1206
Db	2060	GOYGOWAPKLARLHYSGSINAMSTKEPFSWIKVYDLAPMIITHIGIKTQAGRKQFSSYISQ	2119
QY	1207	FIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSGIGKHNIPNPIIARYIRLHPHTHSIRS	1266
Db	2120	FIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSGIGKHNIPNPIIARYIRLHPHTHSIRS	2179
QY	1267	TLRMELMGCDLNSCSPLGNESKXISDAQITASSYFTNPFATWSPSKARLHLOGRNAR	1326

```

Db      2180  TLRMELMGCDLNSCSMPLGWSKAIISDAQITASSYTNMFATWSPSKARLHLQGRNMR 2239
Qy      1327  POWNNEKEWLQVDFQKTMKVTGTTTQGVKSLLTSMVYKFLISSODGHWTLFFQNGKV 1386
Db      2240  POWNNEKEWLQVDFQKTMKVTGTTTQGVKSLLTSMVYKFLISSODGHWTLFFQNGKV 2299
Qy      1387  KVPQGNQDSFTPVVNSLDPLLRTRYLRIRHPQSVVHQAIALRMEVLGCEAODLY 1438
Db      2300  KVPQGNQDSFTPVVNSLDPLLRTRYLRIRHPQSVVHQAIALRMEVLGCEAODLY 2351

RESULT 11
US-10-095-718-4
; Sequence 4, Application US/10095718
; Publication No. US20020131956A1
; GENERAL INFORMATION:
; APPLICANT: Walsh, Christopher
; APPLICANT: Chao, Hengjun
; APPLICANT: Burstein, Haim
; APPLICANT: Lynch, Carmel
; APPLICANT: Stepan, Tony
; APPLICANT: Munson, Keith
; TITLE OF INVENTION: Adeno-Associated Virus Vectors Encoding Factor VIII and
; FILE REFERENCE: Methods of Using the Same
; CURRENT APPLICATION NUMBER: US/10/095,718
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 09/689,430
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/158,780
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1431
; TYPE: PRT
; ORGANISM: canine B-domain deleted factor VIII
US-10-095-718-4

Query Match      85.8%; Score 6599; DB 14; Length 1431;
Best Local Similarity 85.6%; Pred. No. 0;
Matches 1229; Conservative 80; Mismatches 99; Indels 28; Gaps 6;

Qy      1  ATRRYLGAVELSWDMQSD-LGELPVDARFPFPRVPSKSPFNSTVYVYKTLFVEFTVHLF 59
Db      20  ATRKYVIGAVELSWDMQSDLLSALHADTSFSRVPGLSLPTTSVYRKTIVFEFTDDL 79
Qy      60  NIAKPRPPMGLLGPPTQAEVYDTVTITLKNMASHPVSLHAGVSYWASEGAEYDDQTS 119
Db      80  NIAKPRPPMGLLGPPTQAEVYDTVTITLKNMASHPVSLHAGVSYWASEGAEYDDQTS 139
Qy      120  QREKEDDKVFPGGSHTVVQVLKENGPMASDPLCLTYSVLSHVDLVKDLNSGLIGALLVC 179
Db      140  QREKEDDNVTPGSHSTTVVQVLKENGPMASDPPCLTYSVSHVDLVKDLNSGLIGALLVC 199
Qy      180  REGSLAKEKTQTLHKFTILLFAVDEGKSWHSETKNSLSMQDRDAASARAPKMTHTVNGVN 239
Db      200  KESSLAKERQTQLQEFVLLFAVDEGKSWHSETNASITQ-----AEAQELHTINGVYN 253
Qy      240  RSLPGLIGCHRKSVYWHVIGMTTPEVHSIFLEGHTFLVRNHRQASLEISPIFTLTAQTL 299
Db      254  RSLPGLTVCHRSVYWHVIGMTTPEVHSIFLEGHTFLVRNHRQASLEISPIFTLTAQTL 313
Qy      300  LMDLGQFLFCHLSSHQHDGMEAVYKVDSCPEEPQLRKNNEAEDYDDDLTDSMDVVR 359
Db      314  LMDLGQFLFCHPSHQHDGMEAVYKVDSCPEEPQLRKNNEED-KOYDDDLGSDMDVWS 372
Qy      360  FDDNNSPFIQIRSVAKKPKTWVHYIAAEEEDMDYAPLVAPDDRYSKSOYLNNGPQRI 419
Db      373  FDDSSSPFIQIRSVAKKPKTWVHYIAAEEEDMDYAPSGPTPNDRSHKDLNNGPQRI 432
Qy      420  GRKYKVRFMAYTDETFKTRAIQHSGLIGPLLYGEGVDTLLIFRKNQASRPYNIYPHG 479

```

```

Db      433  GKYYKVRFAVYTDFTFKTRAIQYESGILGPELLYGEVGDTLIIIFKNQASRPYNIYPHG 492
Qy      480  ITDVRPLYSRRLPKGVKHLKDPPIILPGEIFKTKWTVTVEDGPTKSDPRCLTRYSSFVNM 539
Db      493  INVVTPLHTRGLPKGVKHLKDPPIILPGEIFKTKWTVTVEDGPTKSDPRCLTRYSSFINL 552
Qy      540  ERDLASGLIGPLLI CYKESVDQRGNQIMSDKRNVLPSVFDENRNSWYLTENTQRFIPNPA 599
Db      553  ERDLASGLIGPLLI CYKESVDQRGNQIMSDKRNVLPSVFDENRNSWYLTENTQRFIPNAD 612
Qy      600  GVOLEDPEFOASNIMHSINGYVFDLSQLSVCLHEVAYWILSIGAOTDFLSVFFSGVTFK 659
Db      613  VVQPHDPFQLSNIMHSINGYVFDNLQSVCLHEVAYWILSIGAOTDFLSVFFSGVTFK 672
Qy      660  HKWVYEDTLTLFPFSGETVFMENPGLMWILGCHNSDFRNRMGTALLKSSCDKNDGYI 719
Db      673  HKWVYEDTLTLFPFSGETVFMENPGLMWILGCHNSDFRNRMGTALLKSSCDNRNDDYI 732
Qy      720  EDSYEDISAYLLSKNAIEPRSFQNPVVKHQRHITRTTLOSDEEDYDDTISVEMK 779
Db      733  EDTYEDIPTPLNENNVIKPRSFQNPVVKHQRHITRTTLOSDEEDYDDTISVEMK 772
Qy      780  KEDFDLYDEDENGSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNRNAGSVPQPKV 839
Db      773  REDFDIYDIENQGLRSFQKTRHYFIAAVERLWDYGMSSSPHVLNRNAGSVDVQPKV 832
Qy      840  VFQETDGSFTQPLRYGELNEHLGLGPIYRAEVEDNIMVTFRNQASRPYSFYSSLSIYE 899
Db      833  VFQETDGSFTQPLRYGELNEHLGLGPIYRAEVEDNIMVTFRNQASRPYSFYSSLSIYD 892
Qy      900  EDORQGAERPKNFVKNETKTYFWKQVHMAPTKDFCKAMAYFSDVLEKDVHSGSLIG 959
Db      893  EDEGQGAERPRKRVNPNETKIYFWKQVHMAPTKDFCKAMAYFSDVLEKDVHSGSLIG 952
Qy      960  PLLVCHTNTLNPAGHGRQVTVQEPALPFTTFDETKSMYFTENMERNCRAPCNIOEPTFK 1019
Db      953  PLLICRSNTLNPAHGRQVTVQEPALPFTTFDETKSMYFTENMERNCRAPCNIOEPTFK 1012
Qy      1020  ENYRFHAINGYIMDTLPLGLVMAQDQIRWYLLSMGNSNENIHSIFSGHVTFRKKEBEYKM 1079
Db      1013  ENYRFHAINGYIMDTLPLGLVMAQDQIRWYLLSMGNSNENIHSIFSGHVTFRKKEBEYKM 1072
Qy      1080  ALYNLPGVPEVTEMPLPSKAGIWRVECLIGEHLAGMSTLFLVYSKQCTPLGWSGHIR 1139
Db      1073  AVYNLPGVPEVTEMPLPSQGIWRVECLIGEHLAGMSTLFLVYSKQCTPLGWSGHIR 1132
Qy      1140  DFOITASGQYQWAPKLARLHYSGSINAMSTKEPFSWIKVDLLAPMIHGIKTQGAOKF 1199
Db      1133  DFOITASGQYQWAPKLARLHYSGSINAMSTKDPFWSIKVDLLAPMIHGIKTQGAOKF 1192
Qy      1200  SSIIYISQFIIMYSLDGKKQTYRGNSTGTILWVFGNVDSGIGHNIFNPPIIARYIRLHP 1259
Db      1193  SSIIYISQFIIMYSLDGKKQTYRGNSTGTILWVFGNVDSGIGHNIFNPPIIARYIRLHP 1252
Qy      1260  THYSIRSTLRMELMGCDLNSCSMPLGWSKAIISDAQITASSYTNMFATWSPSKARLHLQ 1319
Db      1253  THYSIRSTLRMELMGCDLNSCSMPLGWSKAIISDAQITASSYTNMFATWSPSKARLHLQ 1312
Qy      1320  GRSNARPPQVNNPKWLOYDFQKTMKVTGTTTQGVKSLLTSMVYKFLISSODGHWTL 1379
Db      1313  GRTNARPPQVNNPKWLOYDFQKTMKVTGTTTQGVKSLLTSMVYKFLISSODGHWTL 1372
Qy      1380  FFQNGKVKVFGQNDGFTPVVNSLDPLLRTRYLRIRHPQSVVHQAIALRMEVLGCEAO 1435
Db      1373  FLQNGKVKVFGQNDGFTPVVNSLDPLLRTRYLRIRHPQSVVHQAIALRMEVLGCEAO 1428

RESULT 12
US-10-131-510A-39
; Sequence 39, Application US/10131510A
; Publication No. US20030166536A1
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S

```

TITLE OF INVENTION: Modified Factor VIII  
FILE REFERENCE: 75-95J  
CURRENT APPLICATION NUMBER: US/10/131,510A  
PRIOR FILING DATE: 2002-07-10  
PRIOR APPLICATION NUMBER: U.S. 09/315,179  
PRIOR FILING DATE: 1999-05-20  
PRIOR APPLICATION NUMBER: U.S. 09/037,601  
PRIOR FILING DATE: 1998-03-10  
PRIOR APPLICATION NUMBER: U.S. 08/670,707  
PRIOR FILING DATE: 1996-06-26  
PRIOR APPLICATION NUMBER: PCT/US97/11155  
PRIOR FILING DATE: 1997-06-26  
PRIOR APPLICATION NUMBER: PCT/US94/13200  
PRIOR FILING DATE: 1994-11-15  
PRIOR APPLICATION NUMBER: U.S. 08/212,133  
PRIOR FILING DATE: 1994-03-11  
PRIOR APPLICATION NUMBER: U.S. 07/864,004  
PRIOR FILING DATE: 1992-04-07  
NUMBER OF SEQ ID NOS: 40  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 39  
LENGTH: 1443  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: factor VIII lacking  
OTHER INFORMATION: the B domain  
US-10-131-510A-39

Query Match 84.4%; Score 6489; DB 12; Length 1443;  
Best Local Similarity 83.4%; Pred. No. 0;  
Matches 1200; Conservative 104; Mismatches 119; Indels 16; Gaps 3;

1 ATRRYLGAVELSDWYMSQD-LGELPVDARPPRPVPSKSPFWTSVYVYKTLFVETVHLF 59  
20 AIRRYLGAVELSDWYMSQD-LGELPVDARPPRPVPSKSPFWTSVYVYKTLFVETVHLF 79  
60 NTAQRPPMGLGPTIOAEVYDVTITLKNASHPVSLHAGVSVYKASEGAEYDDQTS 119  
80 SVARPPMGLGPTIOAEVYDVTITLKNASHPVSLHAGVSVYKASEGAEYDDQTS 139  
120 QREKEDKVPFGSHTYVQVILKENGPMASDPLCLTYSLYSHVDLVKDLNSGLIGALLVC 179  
140 QREKEDKVPFGSHTYVQVILKENGPMASDPLCLTYSLYSHVDLVKDLNSGLIGALLVC 199  
180 REGSLAKEVQTHLKFILLFAVDECKSWHSETKNSLMQDRDAASARAPKMTYNGYN 239  
200 REGSLAKEVQTHLKFILLFAVDECKSWHSETKNSLMQDRDAASARAPKMTYNGYN 259  
240 RSLPGLIGCHRSVYVHVGMTTPEVHSIFLEGHTFLVRNHRQASLEISPTIFLTAQTL 299  
260 RSLPGLIGCHRSVYVHVGMTTPEVHSIFLEGHTFLVRNHRQASLEISPTIFLTAQTL 319  
300 LMDLGQFLFCHISSHHGMEARVSVCAEPPQLRRKADDER-EDYDNLXDSDMDVVR 378  
320 LMDLGQFLFCHISSHHGMEARVSVCAEPPQLRRKADDER-EDYDNLXDSDMDVVR 398  
360 FDDNDSPTQIRSVAKHPKTHVHYIAAEERDWDYAPLAPDERSYKSOYLNNGPQRI 419  
379 LDGDDVSPFTQIRSVAKHPKTHVHYIAAEERDWDYAPLAPDERSYKSOYLNNGPQRI 438  
420 GRKYKVRFWAYTDETPKTEATQESGILGPLYGEVGDITLIIIFKQASRPYNYPHG 479  
439 GRKYKVRFWAYTDETPKTEATQESGILGPLYGEVGDITLIIIFKQASRPYNYPHG 498  
480 ITDVRPLYSRLPRGVKHLKDFPILPGEIFKYKWTVTVEGPTKSDPRCLTRYSSFVNM 539  
499 ITDVSALHPGRLKGMHLKDMPLPGETFKYKWTVTVEGPTKSDPRCLTRYSSINL 558  
540 ERLASGLIGPLICYKESVDQGNQIMSKRNVILFSPVDENRSLTNIORFLPNPA 599  
559 EKDLASGLIGPLICYKESVDQGNQIMSKRNVILFSPVDENRSLTNIORFLPNPD 618

600 GVQLDEDFQASNMHSINGYVDSLSQVCLHEVAYVWYLLSICAOITDFLSVPSGVTFK 659  
619 GLQPDPEFQASNMHSINGYVDSLSQVCLHEVAYVWYLLSICAOITDFLSVPSGVTFK 678  
660 HKWYEDTLLTLPFSGETVPMSENGPGLWILGCHNSDFRNGMTALLKVSCKNTGDYY 719  
679 HKWYEDTLLTLPFSGETVPMSENGPGLWILGCHNSDFRNGMTALLKVSCKNTGDYY 738  
720 EDSYEDISAYLLSKNNAIEPRSPSONPPVLRKHOREITRTTLQSDQBEIDYDDTISVEMK 779  
739 DNTYEDIPGLLSGKNIEP-----RDISLFTFOPEEDKMDYDDIFSTETK 784  
780 KEDFDIYDEBENQSPRSFQKTRHYFTAAVERLWDYCMSSSPHYLRRAQSGVPOPKV 839  
785 GEDFDIYDEBENQSPRSFQKTRHYFTAAVERLWDYCMSSSPHYLRRAQSGVPOPKV 844  
840 VFOEFTDGSFTQPLRYGELNEHGLLGPYIRAEVEDNIMVTFRQASRPYSFYSSLSIYE 899  
845 VFREFADGSFTQPLRYGELNEHGLLGPYIRAEVEDNIMVTFRQASRPYSFYSSLSIYP 904  
900 EDQAGAEPRKPNFVKNETKTYFWKQVHMAPTKDEFCCKAWAYPSVDLEKOVHSGLIG 959  
905 DDQAGAEPRKPNFVKNETKTYFWKQVHMAPTKDEFCCKAWAYPSVDLEKOVHSGLIG 964  
960 PLLVCHTNTLPAHGRQVTVQEFALFTTIFDETSWYFTENWERNCRAPCNIOMEDPTFK 1019  
965 PLLICRANTLNAHGRQVTVQEFALFTTIFDETSWYFTENWERNCRAPCNIOMEDPTFK 1024  
1020 ENYFPAINGYIMDTLPGLVMAQORIRWYLLSGNSNENIHSIHFSGHVFTVRKKEEYKM 1079  
1025 ENYFPAINGYIMDTLPGLVMAQORIRWYLLSGNSNENIHSIHFSGHVFTVRKKEEYKM 1084  
1080 ALYNYLPGVFTVEMLPSKAGIWRVECLIGHLHAGMSTLFLVYSNKCQTPIGMASGHIR 1139  
1085 AVYNYLPGVFTVEMLPSKAGIWRVECLIGHLHAGMSTLFLVYSNKCQTPIGMASGHIR 1144  
1140 DFOITASQYQGWAPKPLARLHYSGINAWSTKPSNLTVDLLAPMIHGLTKQCARQKF 1199  
1145 DFOITASQYQGWAPKPLARLHYSGINAWSTKPSNLTVDLLAPMIHGLTKQCARQKF 1204  
1200 SSLYISQFIIMYSLDGRNQVYRGNSGTGLVFFGNVDSSGKKNIFNPPIIYIRLHP 1259  
1205 SSLYISQFIIMYSLDGRNQVYRGNSGTGLVFFGNVDSSGKKNIFNPPIIYIRLHP 1264  
1260 THYSIRSTLRMELMGCDLNSCMLPGMOKAISDSQITASSHLSNIFATWSPSQARLHLQ 1319  
1265 THYSIRSTLRMELMGCDLNSCMLPGMOKAISDSQITASSHLSNIFATWSPSQARLHLQ 1324  
1320 GRSNARQVNNPKEMLQVDFQKTKVGTGVTQGVKSLTSMVYKFEFLISSQDGHOWTL 1379  
1325 GRSNARQVNNPKEMLQVDFQKTKVGTGVTQGVKSLTSMVYKFEFLISSQDGHOWTL 1384  
1380 PFQNGKVKVQGNQSDFTPVVNSLDPPLITRYLIRHPSQVSWHQAIALRMEVLGCEADLY 1438  
1385 FLQDGHKVKVQGNQSDFTPVVNSLDPPLITRYLIRHPSQVSWHQAIALRMEVLGCEADLY 1443

RESULT 13  
US-10-187-319-39  
Sequence 39, Application US/10187319  
Publication No. US20030068785A1  
GENERAL INFORMATION:  
APPLICANT: Lollar, John S.  
TITLE OF INVENTION: Hybrid Human/Animal Factor VIII  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Greenlee, Winner and Sullivan, P.C.  
STREET: 5370 Manhattan Circle Suite 201  
CITY: Boulder  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80303  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/187,319  
FILING DATE: 27-Aug-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 09/523,656  
FILING DATE: 2000-03-10  
APPLICATION NUMBER: US 09/037,601  
FILING DATE: 1998-03-10  
APPLICATION NUMBER: WO PCT/US97/11155  
FILING DATE: 1997-06-26  
APPLICATION NUMBER: US 08/670,707  
FILING DATE: 1996-06-26  
ATTORNEY/AGENT INFORMATION:  
NAME: Greenlee, Lorraine L.  
REGISTRATION NUMBER: 27,894  
REFERENCE/DOCKET NUMBER: 75-95K  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 303/499-8080  
TELEFAX: 303/499-8089  
INFORMATION FOR SEQ ID NO: 39:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1443 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 39:  
US-10-187-319-39

Query Match 84.4%; Score 6489; DB 15; Length 1443;  
Best Local Similarity 83.4%; Pred. No. 0;  
Matches 1200; Conservative 104; Mismatches 119; Indels 16; Gaps 3;

QY 1 ATRRYLGAVELSDWYQSD-LGELPVDARPPRPVPSFPNTSVVTKTLFVEFTWHLF 59  
DB 20 AIRRYLGAVELSDWYQSD-LGELPVDARPPRPVPSFPNTSVVTKTLFVEFTWHLF 79

QY 60 NIAKPRPNWGLGPTTQAEVYDVTWTLKNMASHPVSLHGVSYKASEGARYDDOTS 119  
DB 80 SVARPRPNWGLGPTTQAEVYDVTWTLKNMASHPVSLHGVSYKASEGARYDDOTS 139

QY 120 QREKEDDKVPFGSGHTYVWVLKENGPMASDPLCLTYSLSHVDLVKDLNSGLIGALLVC 179  
DB 140 QREKEDDKVPFGSGHTYVWVLKENGPMASDPLCLTYSLSHVDLVKDLNSGLIGALLVC 199

QY 180 REGSLAKETQTLHKFILLFAVDEGKSWHSETKNSLMQORDAASARAPMGHTVNGYVN 239  
DB 200 REGSLAKETQTLHKFILLFAVDEGKSWHSETKNSLMQORDAASARAPMGHTVNGYVN 259

QY 240 RSLPGLIGCHRSYVHVGWGTTPVHSIFLEGHTFLVRHRCASLEISPTFLTAQTL 299  
DB 260 RSLPGLIGCHRSYVHVGWGTTPVHSIFLEGHTFLVRHRCASLEISPTFLTAQTL 319

QY 300 LMDLQGFLLFCHSHSHHGMEAVYKVDSCPEEPQLMKNEAEYDDDLTDSMDVVR 359  
DB 320 LMDLQGFLLFCHSHSHHGMEAVYKVDSCPEEPQLMKNEAEYDDDLTDSMDVVR 378

QY 360 FDDNPSPTQIRSVAKKPKTWVHYIAAEEEDWDYAPLVAPDDRYSKYQYLANGPQRI 419  
DB 379 LDGDDVSPFIQIRSVAKKPKTWVHYIAAEEEDWDYAPLVAPDDRYSKYQYLANGPQRI 438

QY 420 GRKYKVRPMAYTDETPKTRAIQESGILGPLYGEVGTLLIIFKNQASRPNIYPHG 479  
DB 439 GRKYKARFVAYTDETPKTRAIQESGILGPLYGEVGTLLIIFKNQASRPNIYPHG 498

QY 480 ITDVRPLYSRELPGVGHKDFPILPGEIFYKVTWVEDGPKSDRCLTRYYSSTFNM 539  
DB 499 ITDVSALHPGRLKGWHLKMPILPGEIFYKVTWVEDGPKSDRCLTRYYSSTFNM 558

## RESULT 14

US-10-131-510A-37  
Sequence 37, Application US/10131510A  
Publication No. US20030166536A1  
GENERAL INFORMATION:  
APPLICANT: Lollar, John S  
TITLE OF INVENTION: Modified Factor VIII  
FILE REFERENCE: 75-95J  
CURRENT APPLICATION NUMBER: US/10/131,510A  
CURRENT FILING DATE: 2002-07-10  
PRIOR APPLICATION NUMBER: U.S. 09/315,179  
PRIOR FILING DATE: 1999-05-20



;; PRIOR APPLICATION NUMBER: U.S. 09/037,601  
;; PRIOR FILING DATE: 1998-03-10  
;; PRIOR APPLICATION NUMBER: U.S. 08/670,707  
;; PRIOR FILING DATE: 1996-06-26  
;; PRIOR APPLICATION NUMBER: PCT/US97/11155  
;; PRIOR FILING DATE: 1997-06-26  
;; PRIOR APPLICATION NUMBER: PCT/US94/13200  
;; PRIOR FILING DATE: 1994-11-15  
;; PRIOR APPLICATION NUMBER: U.S. 08/212,133  
;; PRIOR FILING DATE: 1994-03-11  
;; PRIOR APPLICATION NUMBER: U.S. 07/864,004  
;; PRIOR FILING DATE: 1992-04-07  
;; NUMBER OF SEQ ID NOS: 40  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 37  
;; LENGTH: 2133  
;; TYPE: PRT  
;; ORGANISM: Sus scrofa  
US-10-131-510A-37

Query Match 81.0%; Score 6227; DB 12; Length 2133;

Best Local Similarity 57.3%; Pred. No. 0;  
Matches 1212; Conservative 106; Mismatches 119; Indels 678; Gaps 3;

QY 1 ATRRYLGAVELSDWIMQSD-IGLEFVDARPPRPVKSPFNNTSVYKTLFVFEVTHLF 59  
DB 20 AIRRYLGAVELSDWYRQSELLRELHVDTRFATAPGALPLGSPVLYKKTVFVEVTDQLF 79  
QY 60 NIAKRRPPMGLLGTIOAEVYDVTIVILKNMASHPVSLHAYGVSNWKAEGAEVDDOTS 119  
DB 80 SVAREPPMGLLGTIOAEVYDVTIVILKNMASHPVSLHAYGVSNWKAEGAEVDDOTS 139  
QY 120 QREKEDDKVPFGSGSHYVMOVLKENGPMASDPLCLTYSLSHVDLVKDLNSGLIGALLVC 179  
DB 140 QREKEDDKVLPKGSQTYVMQVLKENGPMASDPLCLTYSLSHVDLVKDLNSGLIGALLVC 199  
QY 180 REGSLAKEKTQTLHFKFILLFAVDFDEKSWHSETKNSLMQDRDAASARAPKMPKMTVNGYN 239  
DB 200 REGSLTRETONLHERVLLFAVDFDEKSWHSEARNDSTWTRAMPAPAPAPAMHTVNGYN 259  
QY 240 RSLPLGLIGHRSVYVHVIGMGTTPPEVHSIFLEGHTFLVRNHRQASLEISPTFLTAQTL 299  
DB 260 RSLPLGLIGHRSVYVHVIGMGTTPPEVHSIFLEGHTFLVRNHRQASLEISPTFLTAQTF 319  
QY 300 LMDLGOFLFCHLSSHQHGMENYKVDSCPEPQLRMKNNEAEYDDDLTDSMDVVR 359  
DB 320 LMDLGOFLFCHLSSHHGMEAHVRVSCAEPEQLRRKADBE-EDYDDNLVDSMDVVR 378  
QY 360 FDDNSPSTQIRSVAKGHPKTVVHYIAAEEEDWDYAPLVAPDPRSYSKYQVYLNNGPQRI 419  
DB 379 LDGDDVSPFIQIRSVAKGHPKTVVHYIAAEEEDWDYAPLVAPDPRSYSKYQVYLNNGPQRI 438  
QY 420 GRKYKVRWYATDETFKTRATQHSGLILGPLYGEGVDTLLIIFKNQASRPYNYLPHG 479  
DB 439 GRKYKARFVAYDVTFKTRATQHSGLILGPLYGEGVDTLLIIFKNQASRPYNYLPHG 498  
QY 480 ITDVRPLSRLLPKGVKHLKDFPILGELIFKYKWTVTVEDGPKSPRCLTRYSSFVNM 539  
DB 499 ITDVSALHPGRLLKWKHLKDFPILGELIFKYKWTVTVEDGPKSPRCLTRYSSINL 558  
QY 540 ERDLASGLIGPLLIICYKESVDQRNQIMSDKRVNLLFSVFDENRWSYLTFENIORFLPNPA 599  
DB 559 EKDLASGLIGPLLIICYKESVDQRNQIMSDKRVNLLFSVFDENRWSYLTFENIORFLPNPD 618  
QY 600 GVQLEDEPQASNMHSINGYVFDLSQLSVCLHEVAYWYILSTGAQTDFLSVFSGYTFK 659  
DB 619 GLQPDQEPQASNMHSINGYVFDLSQLSVCLHEVAYWYILSTGAQTDFLSVFSGYTFK 678  
QY 660 HKMUYEDTLTLFPESGTFVMSMENPGLWILGCHNSDFRNRGMTALLKVSSCDKNTGDYY 719  
DB 679 HKMUYEDTLTLFPESGTFVMSMENPGLWILGCHNSDFRNRGMTALLKVSSCDKNTGDYY 738  
QY 720 EDSVEDISAYLLSKNNAIEPRFSQN----- 745

Db 739 DNTYEDIPGLLSGKNVIEPRSPAQRPPPSASQKQPTTTSDEDDVELDPQSGERTQAL 798  
QY 746 ----- 745  
Db 799 EELSVPGDGMGLLQGNPAPHGSSSSDLQEARNEADDYLPGARERTAPSAARLRDELH 858  
QY 746 ----- 745  
Db 859 HSAERVLTPEPEKELKKLDSKMSSSSDLLKTSPTIPSDTILSAETERTHSLGPPHPQVNF 918  
QY 746 ----- 745  
Db 919 SOLGAILVGNSSHFICAGVPLGSTEEDHESSLGENVSPVESDGIPEKRAHGASLTOKD 978  
QY 746 ----- 745  
Db 979 DVLFKVNSILVKTNKARVILKTNKRIHIDDAALLTENRASATFMDKNTTASGLNHVSNWI 1038  
QY 746 ----- 745  
Db 1039 KGPLGNPLSSERGSPPELLTSSGSGKSVKQSSGQGRIRVAVEEBELSKGKEMMLPNSE 1098  
QY 746 ----- 745  
Db 1099 LTFLTNSADVQGNTHSGGKKSREMERREKLVOEKVDLPQVYTATGTYKFLNIFHQST 1158  
QY 746 ----- 745  
Db 1159 EPSVEGPDGSHAPVPQDSRSLNDSABRAETHIAHFAISAREEAPLEAPNGRTGPPRSV 1218  
QY 746 ----- 745  
Db 1219 PRRVKQSLKQIRLPLEEIKPERGVVLTATSTWSESSPILQAKNNLSLPLTLEMAGG 1278  
QY 746 ----- 745  
Db 1279 QKISALGSAAGPLASGLKELAVLSSAGLSEASGABFLPKVRVHRDILLPQKTSNVSC 1338  
QY 746 ----- 745  
Db 1339 AHGDLQGBIFLQKTGFPVNLKNVNRPGRTPTSKLLGPPMPKEWSELESKPSKSTALTKDII 1398  
QY 746 ----- 745  
Db 1399 SLPLDRHESNHSIAAKNEQAETQREAAATWQKGGPGLCAPKPPVLRHRQDILSLPTQP 1458  
QY 764 DQEEIDYDDTISVEMKKEDFDIYDEENQSPSPFOKTRHYFIAAVERLWDYGMSSSPHV 823  
Db 1459 EEDKMDYDDITFSTKGGEDFDIYGEDENQDPRSPQKTRHYFIAAVERLWDYGMSSSPHV 1518  
QY 824 LENRAQSGSVPOFKVVFQEFDTGSGFTQPLRGELNEHLLGLPGYIRAEVEDNIMVTFN 883  
Db 1519 LENRAQSGSVPOFKVVFQEFDTGSGFTQPLRGELNEHLLGLPGYIRAEVEDNIMVTFN 1578  
QY 884 QASRPYSFSSLSIIVEEDORQCAEPKPKFVKNETKTFKVOHMAFKDPDCKWAY 943  
Db 1579 QASRPYSFSSLSIIVEEDORQCAEPKPKFVKNETKTFKVOHMAFKDPDCKWAY 1638  
QY 944 FSDVLEKDVHSLGLPLLCVCHTNTLNPAGHQVTVQEPALFTTIFDETKSWYFENMER 1003  
Db 1639 FSDVLEKDVHSLGLPLLCVCHTNTLNPAGHQVTVQEPALFTTIFDETKSWYFENMER 1698  
QY 1004 NCRAPCNIOEDPTPKENYRFAHNGYIMDTLPLGLVMAQDQRIRWYLLSMGSMENIHSIH 1063  
Db 1699 NCRAPCNIOEDPTPKENYRFAHNGYIMDTLPLGLVMAQDQRIRWYLLSMGSMENIHSIH 1758  
QY 1064 FSGHVTFVKKEBYKMAVNLNYPGVFETVEMLPSTKAGIWRVBECLTGEHLHAGMSTFLVY 1123  
Db 1759 FSGHVTFVKKEBYKMAVNLNYPGVFETVEMLPSTKAGIWRVBECLTGEHLHAGMSTFLVY 1818  
QY 1124 SNKQOTPLWASGHTEFQITASGOYQWAPKLARLHYSGINASSTKPPSWIKVDLLA 1183

Db 1819 SKECAPLGMASGRIRDRQITASGOYQOWAPKLARLHYSGSINAWSTKDPHSWIKVDLLA 1878  
 Qy 1184 PMIHGINTQARQKFSLLYISQFIIMYSLDGKKWQTYRGNSTGFLMWFFGQVNDSSGIGH 1243  
 Db 1879 PMIHGINTQARQKFSLLYISQFIIMYSLDGKNWQSYRGNSTGFLMWFFGQVNDASGIGH 1938  
 Qy 1244 NIFNPPITARYLRHPTHTYSIRSTRMELMCGDLNCSMPILGMSKASDAQITASSYFT 1303  
 Db 1939 NIFNPPITARYLRHPTHTYSIRSTRMELMCGDLNCSMPILGMSKASDAQITASSHLS 1998  
 Qy 1304 NMFATWSPSKARLHQGRSNARPOVNNPKEMQVDFQTKMKVTGVTTOGVKSLITSMYV 1363  
 Db 1999 NIFATWSPQARLHQGRSNARPRVSSAEWLQVDLQTKVTVGITTQGVKSLSSMYV 2058  
 Qy 1364 KEFLISSQDGHQMTLFPQNGKVKVFCQGNQDSFTFPVNSLPPPLTRYLRHPOSWHQI 1423  
 Db 2059 KEFLVSSQDGRRTLFLQDGHKTVFCQGNQDSSTEVVNALDPLFLTRYLRHPTSWAQHI 2118  
 Qy 1424 ALRMEVLGCEAQDLY 1438  
 Db 2119 ALRLEVLCGEAQDLY 2133

## RESULT 15

US-10-187-319-37  
 ; Sequence 37, Application US/10187319  
 ; Publication No. US20030068785A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lollar, John S.  
 ; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII  
 ; NUMBER OF SEQUENCES: 40  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.  
 ; STREET: 5370 Manhattan Circle Suite 201  
 ; CITY: Boulder  
 ; STATE: Colorado  
 ; COUNTRY: USA  
 ; ZIP: 80303  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/10/187,319  
 ; FILING DATE: 27-Aug-2002  
 ; CLASSIFICATION: <Unknown>  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 09/523,656  
 ; FILING DATE: 2000-03-10  
 ; APPLICATION NUMBER: US 09/037,601  
 ; FILING DATE: 1998-03-10  
 ; APPLICATION NUMBER: WO PCT/US97/11155  
 ; FILING DATE: 1997-06-26  
 ; APPLICATION NUMBER: US 08/670,707  
 ; FILING DATE: 1996-06-26  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Greenlee, Lorraine L.  
 ; REGISTRATION NUMBER: 27,894  
 ; REFERENCE/DOCKET NUMBER: 75-95K  
 ; TELEPHONE: 303/499-8080  
 ; TELEFAX: 303/499-8089  
 ; INFORMATION FOR SEQ ID NO: 37:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2133 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 37:  
 US-10-187-319-37

Query Match 81.0%; Score 6227; DB 15; Length 2133;

Best Local Similarity 57.3%; Pred. No. 0;  
 Matches 1212; Conservative 106; Mismatches 119; Indels 678; Gaps 3;  
 Qy 1 ATRYYILGAVELSDYMQSD-LGELPVDARPPRPVPSFPENTSVVYKTLFVFTVHLF 59  
 Db 20 AIRYYILGAVELSDYMQSLLRELHVDTRFPAPGALPLGPSVLYKTVFVFTDOLF 79  
 Qy 60 NIAKPRPPMGLLPTTQAEVYDVTITLKNMASHPSVLSHAGVSYVWASGAYDDOTS 119  
 Db 80 SVARPPRPMMGLLPTTQAEVYDVTITLKNMASHPSVLSHAGVSYVWASGAYDDOTS 139  
 Qy 120 QREKDDKVPFGGSHYVQVVLKENGPMASDPLCTYSYLSHVDLVKDLNSGLGALLVC 179  
 Db 140 QREKDDKVPFGGSHYVQVVLKENGPMASDPLCTYSYLSHVDLVKDLNSGLGALLVC 199  
 Qy 180 REGSLAKEKTKTLHKFILLFAVFDGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVN 239  
 Db 200 REGSLTREIRIQNLHEFVLLFAVFDGKSWHSEARNDSTRAMDPAPARAQPAHMTVNGYVN 259  
 Qy 240 RSLPGLIGCHRKSVYVHVGMTTPEVHSIFLEGHTFLVRNHRQASLEISPIITLTAQTL 299  
 Db 260 RSLPGLIGCHRKSVYVHVGMTTPEVHSIFLEGHTFLVRNHRQASLEISPIITLTAQTF 319  
 Qy 300 LMDLGQFLLFCHISSHODGMEAYVKYDSCPEERPOLRMKNNEAEYDDDLTDSMDVVR 359  
 Db 320 LMDLGQFLLFCHISSHODGMEAYVKYDSCPEERPOLRMKNNEAEYDDDLTDSMDVVR 378  
 Qy 360 FDDNDSPSFIQIRSVAKKHPKTVHYTAAEEDMDYAPLVLPADDRSVKSOYLNNGPQRI 419  
 Db 379 LDGDDVSPFIQIRSVAKKHPKTVHYTAAEEDMDYAPLVLPADDRSVKSOYLNNGPQRI 438  
 Qy 420 GRKYKVRPMAYTDTFTKTRAIQHSIGILGPLLYGEVGDTLIIIFKNQASRPYNIYPHG 479  
 Db 439 GRKYKARFVAYTDTFTKTRAIQHSIGILGPLLYGEVGDTLIIIFKNQASRPYNIYPHG 498  
 Qy 480 ITDYRPLYSRRLPKGVKHLKDFPILPGEIIFYKKTIVTVEDGPTKSDPRCLTRYVSSFNVM 539  
 Db 499 ITDVSALHPGLLKGWKLKDMPTLPGETFYKKTIVTVEDGPTKSDPRCLTRYVSSFNVM 558  
 Qy 540 ERDLASGLIGPLLI CYKESVDORGNQIMSDKRNVLFSVFDENRSWYLTENIQFLPNPA 599  
 Db 559 EKDLASGLIGPLLI CYKESVDORGNQIMSDKRNVLFSVFDENRSWYLTENIQFLPNPD 618  
 Qy 600 GVQLEDPEFQASNTMHSINGYVFDLSQLSVCLHEVAYWYIISIGAQDTDFLSVFFSGYTFK 659  
 Db 619 GLQPDPEFQASNTMHSINGYVFDLSQLSVCLHEVAYWYIISIGAQDTDFLSVFFSGYTFK 678  
 Qy 660 HKMYVEDTLTLFPFSGETVFMENPGLWILGCHNSDFRNRGTMALLKVSCKDNTGDIY 719  
 Db 679 HKMYVEDTLTLFPFSGETVFMENPGLWILGCHNSDFRNRGTMALLKVSCKDNTGDIY 738  
 Qy 720 EDSYEDISAVLLSKNNAIEPRSFQSN----- 745  
 Db 739 DNTYEDIPGILLSGKNVIEPRSFQSNRPPSASOKQQTITSPEDDVELDPQSGERTQAL 798  
 Qy 746 ----- 745  
 Db 799 EELSVPDGSMLLGQNPAPHGSSSSDLQEARNEADDYLPGARERTAPSAARLPELH 858  
 Qy 746 ----- 745  
 Db 859 HSABRVLTPEPEKELKLDKMSSSSLLKTSPTIPSDTLTAETERTHSLGPHHPQVNF 918  
 Qy 746 ----- 745  
 Db 919 SQLGAILVGNSSHFIAGVPLGTEBDEHSSLGENVSPVESDGI FEKERAHGASLTAKD 978  
 Qy 746 ----- 745  
 Db 979 DVLKVNISLVKTNKARVYLKTNRKIHIDDAALLTENPASATFMDKNTTASGLNHVSNWI 1038  
 Qy 746 ----- 745

Search completed: December 9, 2003, 17:01:37  
Job time : 89 secs

Db 1039 KPGLKNPLSSRGSPPELLTSSGSGKSVKQSSGQGRIRVAVBEEELSKGKEMMLPNSE 1098  
QY 746 ----- 745  
Db 1099 LTFLLNSADVQNDTHSQGKKSREMERREKLVBQKVDLPQVYTATGKTNFLRNFHOST 1158  
QY 746 ----- 745  
Db 1159 EPSVEFGDGHAPVQDSRSLNDSABRAETHIAHFSAIRERAPLEAPGNRTGPGPSAV 1218  
QY 746 ----- 745  
Db 1219 PRVKQSLKQIRLPLEEIKPERGVVLNATSTRWSSSPILQGAKRNNLSLPFLTLEMAGG 1278  
QY 746 ----- 745  
Db 1279 QGKISALGKSAAGPLASGLKAVLSSAGLSEASGKAFLPKVRVHREDLLPQKTSNVSC 1338  
QY 746 ----- 745  
Db 1339 AHGDLQOEIPLQKTRGPVNLKVNRPGRTPSKLLGPPMPKEWESLEKSPKSTALRTKDII 1398  
QY 746 -----PPVLKRHOREITRTLOS 763  
Db 1399 SLPLDRHESNHSIAAKNEGQAEQREAAWTKQGGPGLCAPKPPVLRHRRQDISLPTFOP 1458  
QY 764 DOBEIDYDITISVEMKEDPDIYDEBENQSPRSFQKTRHYFIAAVERLMDYGMSSSPHV 823  
Db 1459 EEDKMDYDIFSTETKGEDPDIYGEDENQDPRSFOKTRHYFTAAVEQLWDYGMSPRA 1518  
QY 824 LRNRAQSGSVQPKVKVFOEFTDGSFTOPLYRGELNEHLGLLGPYIRAEVEDNIMVTFRN 883  
Db 1519 LRNRAQGEVPRFKKVVREFADGFTQPSYRGELNKLGLLGPYIRAEVEDNIMVTFKN 1578  
QY 884 QASRPYSFSSLSIYBEDQQAEPKRNFKVKNETKTYFMKVQHMAPTKDEDFCKAWAY 943  
Db 1579 QASRPYSFSSLSIYDQQAEPKRNFKVKNETKTYFMKVQHMAPTKDEDFCKAWAY 1638  
QY 944 FSDVDLEKDVHSLIGLPLLVCHNTNLPALGRQVTVQEFALFTTIDETKSWYFTENMER 1003  
Db 1639 FSDVDLEKDVHSLIGLPLLVCRANTLNAHGRQVTVQEFALFTTIDETKSWYFTENVER 1698  
QY 1004 NCRAPCNIQMEDPTFKENYFHAINGYIMDTPLGLVMAQDQRIRWYLLSMGSENENIHSIH 1063  
Db 1699 NCRAPCHLOMEDPTLKENYFHAINGYIMDTPLGLVMAQDQRIRWYLLSMGSENENIHSIH 1758  
QY 1064 FSGHFTVRKKEEKYKALYNLYPGVFETVEMLPKAGIMRVECLIGEHLHAGMSTFLVY 1123  
Db 1759 FSGHVSVRKKEEKYKAVNLYPGVFETVEMLPKAGIMRVECLIGEHLHAGMSTFLVY 1818  
QY 1124 SNKQOTPLGMAASHIRDFQITAGSQYQGWAPKLARLHYSINNAWSTKEPFSWIKVDLLA 1183  
Db 1819 SKEQAPLGMAASGRIRDFQITAGSQYQGWAPKLARLHYSINNAWSTKDPHSMIKVDLLA 1878  
QY 1184 PMITHGKTGAKOKFSSLYISQFIIMYSLDGKQWTVRGSTCTLMVFFGNDVSSGIKH 1243  
Db 1879 PMITHGIMTOGAROKFSSLYISQFIIMYSLDGKQWTVRGSTCTLMVFFGNDVSSGIKH 1938  
QY 1244 NIFNPPIIARYIRLHPHYTHYSIRSTLRMELMGCDLNSCMPLGMQNKAISSDSQITASSHLS 1303  
Db 1939 NIFNPPIVARYIRLHPHYTHYSIRSTLRMELMGCDLNSCMPLGMQNKAISSDSQITASSHLS 1998  
QY 1304 NMFATWSPKARLHQGSNAWRPQVNNPKEWLQVDFOKTMKVTVGTVTQGVKSLTSMYV 1363  
Db 1999 NIFATWSPSARLHQGRNAWRPRVSSABEWLQVDLOKTVKVTGTVTQGVKSLTSMYV 2058  
QY 1364 KEFLISSQDGHQWTLFPQNGKVKVQCNQDSFTPVNNSLDPPLLTRYLRIHPQSWVHOI 1423  
Db 2059 KEFLVSSQDGRWTLFTQDGHKTVFQCNQDSSTFPVNALDPPLFLTRYLRIHPISWAOHI 2118  
QY 1424 ALRMEVLGCEAQDLY 1438  
Db 2119 ALRLEVIGCEAQDLY 2133

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 9, 2003, 16:51:16 ; Search time 29 Seconds  
(without alignments)  
4768.640 Million cell updates/sec

Title: US-10-006-091-1  
Perfect score: 7691  
Sequence: 1 ATRYYLGAVELSWDMQSD.....WVHQIAURMEVLGCEAQLY 1438

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 76:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7227	94.0	2351	1 EZHU	coagulation factor
2	6227	81.0	2133	2 T42763	coagulation factor
3	6192	80.5	2319	2 A47004	coagulation factor
4	2390.5	31.1	2224	1 KFHUS	coagulation factor
5	2372	30.8	2211	1 KFBOS	coagulation factor
6	2369	30.8	2183	2 T42764	coagulation factor
7	1731.5	22.5	1069	1 KURHU	ferroxidase (BC 1.
8	1687.5	21.9	1059	1 A35210	ferroxidase (BC 1.
9	1104	14.4	216	2 A44258	factor VIII-associ
10	663	8.6	427	2 JC4915	ags protein precu
11	657	8.5	463	1 A36479	milk fat globule m
12	650	8.5	409	2 T11743	pp47 protein - pig
13	635	8.3	401	2 S65138	glycoprotein antig
14	635	8.3	427	2 S74211	PAS-6/7 protein pr
15	543	7.1	869	2 A25945	coagulation factor
16	443	5.8	927	1 JQ9548	A5 antigen precurs
17	424.5	5.5	218	2 A47285	milk fat globule p
18	306.5	4.0	3133	2 S52093	hemocytin - silkw
19	262	3.4	845	2 JC5256	adipocyte transcri
20	221	2.9	719	2 S51739	transcription repr
21	211	2.7	1283	2 T13799	neurexin IV - frui
22	207.5	2.7	737	2 T31349	hypothetical prote
23	206.5	2.7	737	2 T15615	hypothetical prote
24	193	2.5	913	2 A48280	receptor tyrosine
25	192	2.5	876	2 A49508	protein-tyrosine k
26	185.5	2.4	855	2 S42621	protein-tyrosine k
27	185.5	2.4	910	2 A53137	tyrosine kinase re
28	173.5	2.3	819	2 I48859	tyro 10 receptor k
29	171	2.2	1381	2 T31083	paranodin - rat

30	171	2.2	1385	2 T14158	neurexin IV - mous
31	149.5	1.9	578	2 S66353	L-ascorbate oxidas
32	149	1.9	1883	2 G82875	hypothetical prote
33	148	1.9	622	2 S62580	probable multicopp
34	141.5	1.8	1196	2 S46430	botulinum neurotox
35	140.5	1.8	1196	2 J01467	toxin, nontoxic co
36	140.5	1.8	1217	2 T18209	mxg protein - Bac
37	139.5	1.8	2925	2 T00133	RNA-directed RNA p
38	139	1.8	1193	2 JC4901	nontoxic-nonhemagg
39	137.5	1.8	580	2 F84828	probable laccase (
40	137.5	1.8	903	2 F82080	preprotein translo
41	136.5	1.8	2657	2 T18497	hypothetical prote
42	135	1.8	343	2 T35030	probable copper ox
43	134.5	1.7	791	2 T16031	hypothetical prote
44	134.5	1.7	1132	2 H82887	hypothetical prote
45	133.5	1.7	567	2 T44928	L-ascorbate oxidas

ALIGNMENTS

coagulation factor VIII precursor [validated] - human  
N:Alternate names: antihemophilic factor A; coagulation factor VIIIc; procoagulant co  
C:Species: Homo sapiens (man)  
C>Date: 28-Aug-1985 #sequence revision 28-Aug-1985 #text change 08-Dec-2000  
C:Accession: I54318; A00525; I58059; A23584; A26174; A42348; A43986; S63527; S66445;  
R:Gitschier, J.; Wood, W.I.  
Hum. Mol. Genet. 1, 199-200, 1992  
A>Title: Sequence of the exon-containing regions of the human factor VIII gene.  
A:Reference number: I54318; MUID:93265012; PMID:1303178  
A:Accession: I54318  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1921,'S',1923-2351 <RES>  
A:Cross-references: GB:M8648; NID:G182381; PIDN:AAA52420.1; PID:G182383  
R:Wood, W.I.; Capon, D.J.; Simonsen, C.C.; Baton, D.L.; Gitschier, J.; Keyt, B.; Seeb  
Nature 312, 330-337, 1984  
A>Title: Expression of active human factor VIII from recombinant DNA clones.  
A:Reference number: A00525; MUID:85061548; PMID:6438526  
A:Accession: A00525  
A:Molecule type: mRNA  
A:Residues: 1-2351 <WOO>  
R:Toole, J.J.; Knopf, J.L.; Wozney, J.M.; Sultzman, L.A.; Buecker, J.L.; Pittman, D.D  
s, D.N.; Hewick, R.M.  
Nature 312, 342-347, 1984  
A>Title: Molecular cloning of a cDNA encoding human antihemophilic factor.  
A:Reference number: I58059; MUID:85061550; PMID:6438528  
A:Accession: I58059  
A>Status: nucleic acid sequence not shown; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-74,'V',76-1259,'E',1261-2351 <RE2>  
A:Cross-references: GB:K01740; NID:G182802; PIDN:AAA52484.1; PID:G182803  
R:Truett, M.A.; Blacher, R.; Burke, R.L.; Caput, D.; Chu, C.; Dina, D.; Hartog, K.; K  
.B.; Randolph, A.; Urdea, M.S.; Valenzuela, P.; Dahl, H.H.; Favalaro, J.; Hansen, J.;  
DNA 4, 333-349, 1985  
A>Title: Characterization of the polypeptide composition of human factor VIII:C and t  
A:Reference number: A23584; MUID:86081164; PMID:3933400  
A:Accession: A23584  
A:Molecule type: mRNA  
A:Residues: 1-2351 <TRU>  
A:Cross-references: GB:M14113; NID:G182817; PIDN:AAA52485.1; PID:G182818  
R:Eaton, D.; Rodriguez, H.; Vehar, G.A.  
Biochemistry 25, 505-512, 1986  
A>Title: Proteolytic processing of human factor VIII. Correlation of specific cleavag  
ity.  
A:Reference number: A26174; MUID:86159740; PMID:3082357  
A:Accession: A26174  
A:Molecule type: protein  
A:Residues: 20-36;392-399,'X',401-402;1668-1678;1709-1722,'D',1723-1725;1741-1755 <EA  
R:Pittman, D.D.; Wang, J.H.; Kaufman, R.J.

Biochemistry 31, 3315-3325, 1992  
 A:Title: Identification and functional importance of tyrosine sulfate residues within re  
 A:Reference number: A42348; PMID:92207952; PMID:155416  
 A:Accession: A42348  
 A:Molecule type: protein  
 A:Residues: 20-36;356-371;392-408;582-594;1668-1669,'X',1671;1672-1692;1693-1708;1709-17  
 A:Experimental source: recombinant material from Chinese hamster ovary cells  
 A:Note: sequence extracted from NCBI backbone and corrected to correspond with the publi  
 R:Pay, P.J.; Smudzin, T.M.  
 J. Biol. Chem. 264, 14005-14010, 1989  
 A:Title: Intersubunit fluorescence energy transfer in human factor VIII.  
 A:Reference number: A43986; PMID:89340500; PMID:2503509  
 A:Accession: A43986  
 A:Molecule type: protein  
 A:Residues: 'X',517-523;1853-1860,'X',1862-1864,'X',1866 <FAY>  
 R:Leyste, A.; van Schijndel, H.B.; Niehrs, C.; Huttner, W.B.; Verbeet, M.P.; Mertens, K.;  
 J. Biol. Chem. 266, 740-746, 1991  
 A:Title: Sulfation of Tyr(1680) of human blood coagulation factor VIII is essential for  
 A:Reference number: A56109; PMID:91093266; PMID:1898735  
 A:Contents: annotation; sulfation  
 R:Gitschler, J.; Wood, W.I.; Goralka, T.M.; Wion, K.L.; Chen, E.Y.; Baton, D.H.; Vehar,  
 Nature 312, 326-330, 1984  
 A:Title: Characterization of the human factor VIII gene.  
 A:Reference number: A56196; PMID:85061547; PMID:6438525  
 A:Contents: annotation; introns  
 R:McMullen, B.A.; Fujikawa, K.; Davie, E.W.; Hedner, U.; Ezban, M.  
 Protein Sci. 4, 740-746, 1995  
 A:Title: Locations of disulfide bonds and free cysteines in the heavy and light chains o  
 A:Reference number: A56216; PMID:95338127; PMID:7613471  
 A:Contents: annotation; disulfide bonds  
 A:Note: 329-Cys, 711-Cys, and 2019-Cys were shown to have free sulphydryls  
 R:Kjalker, M.; Hedning, A.; Talbo, G.; Persson, E.; Thomsen, J.; Ezban, M.  
 Eur. J. Biochem. 234, 773-779, 1995  
 A:Title: Amino acid residues 721-729 are required for full factor VIII activity.  
 A:Reference number: S63527; PMID:96163459; PMID:8575434  
 A:Accession: S63527  
 A:Molecule type: protein  
 A:Residues: 733-752;753-759 <KJA>  
 R:Lind, P.; Larsson, K.; Spira, J.; Sydow-Baekman, M.; Almstedt, A.; Gray, E.; Sandberg  
 Eur. J. Biochem. 232, 19-27, 1995  
 A:Title: Novel forms of B-domain-deleted recombinant factor VIII molecules. Construction  
 A:Reference number: S66445; PMID:96048024; PMID:7556150  
 A:Accession: S66445  
 A>Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 1668-1685 <LIN>  
 C:Comment: Factor VIII is activated by factor Xa and thrombin, but prolonged exposure ph  
 C:Genetics:  
 A:Gene: GDB:F8C  
 A:Cross-references: GDB:119124; OMIM:306700  
 A:Map position: Xq28-Xq28  
 A:Introns: 48/2; 89/1; 130/1; 201/1; 224/1; 263/1; 337/1; 424/2; 481/3; 513/1; 584/3; 63  
 C:Function:  
 A:Description: acts as a cofactor, with calcium and phospholipid, for the factor IXa pro  
 A:Pathway: blood coagulation  
 C:Superfamily: coagulation factor VIII; discoidin I amino-terminal homology; ferroxidase  
 C:Keywords: acute phase; blood coagulation; duplication; glycoprotein; hemophilia A; pla  
 F:1-19/Domain: signal sequence #status predicted <SIG>  
 F:20-2351/Product: coagulation factor VIII #status experimental <MAT>  
 F:20-740/Product: coagulation factor VIIIA heavy chain #status experimental <ACH>  
 F:20-356/Domain: A1 <DA1>  
 F:23-348/Domain: ferroxidase repeat homology <FO1>  
 F:392-759/Domain: A2 <DA2>  
 F:402-730/Domain: ferroxidase repeat homology <FO2>  
 F:760-1667/Domain: B <DBO>  
 F:1668-2351/Product: coagulation factor VIIIA light chain #status experimental <ACL>  
 F:1709-2038/Domain: A3 <DA3>  
 F:1716-2038/Domain: ferroxidase repeat homology <FO3>  
 F:2039-2191/Domain: C1 <DC1>  
 F:2039-2188/Domain: discoidin I amino-terminal homology <DN1>  
 F:2192-2351/Domain: C2 <DC2>  
 F:2192-2345/Domain: discoidin I amino-terminal homology <DN2>  
 F:60,258,601,776,803,847,919,962,982,1020,1024,1074,1085,1204,1274,1278,1301,1319,1403,1

F:172-198,267-348,547-573,649-730,1851-1877,1918-1922,2040-2188/disulfide bonds: #sta  
 F:355-356/Cleavage site: Arg-Met (coagulation factor Xa, protein C) #status predicted  
 F:365,737,738,742,1683,1699/Binding site: sulfatase (Tyr) #status experiment  
 F:391-392/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experiment  
 F:414,426/Binding site: sulfate (Tyr) (covalent) #status predicted  
 F:759-760/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experiment  
 F:1667-1668/Cleavage site: Arg-Glu (unidentified proteinase) #status experimental  
 F:1708-1709/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experime  
 F:1740-1741/Cleavage site: Arg-Ala (coagulation factor Xa) #status experimental  
 F:2193-2345/disulfide bonds: #status predicted

Query Match 94.0%; Score 7227; DB 1; Length 2351;  
 Best Local Similarity 61.6%; Pred. No. 0;  
 Matches 1437; Conservative 0; Mismatches 1; Indels 894; Gaps 1;

QY	1	ATRRYILGAVELSDWYMSQDLGELPVDARFPVRPKSPFFNTSVYKTLFVEFTVHLFN	60
DB	20	ATRRYILGAVELSDWYMSQDLGELPVDARFPVRPKSPFFNTSVYKTLFVEFTVHLFN	79
QY	61	IAKPRPPWMLLGPTIOAEVDTVTITKNMASHPVSLHVGVSVMKASEGAEYDDQTSQ	120
DB	80	IAKPRPPWMLLGPTIOAEVDTVTITKNMASHPVSLHVGVSVMKASEGAEYDDQTSQ	139
QY	121	REKEDDKVFPGGSHYTVQVLKENGPMASDPCLTYSYLSHVDLVKDLNSGLIGALLVCR	180
DB	140	REKEDDKVFPGGSHYTVQVLKENGPMASDPCLTYSYLSHVDLVKDLNSGLIGALLVCR	199
QY	181	EGSLAKEKTTQTLHKILLFAVDEGKSHSETKNSLMQDRDAASARAWPKMTVGVNR	240
DB	200	EGSLAKEKTTQTLHKILLFAVDEGKSHSETKNSLMQDRDAASARAWPKMTVGVNR	259
QY	241	SUPGLIGCHRKSVYWHVIGMTTPEVHSIFLGHTFLVNRHQASLEISPIFLTAQTLL	300
DB	260	SUPGLIGCHRKSVYWHVIGMTTPEVHSIFLGHTFLVNRHQASLEISPIFLTAQTLL	319
QY	301	MDLGQFLFLFCHSSHQHDGMEAYVYVDSCEPSPQLMKNNRBAEDYDDLTDSEMDVVRP	360
DB	320	MDLGQFLFLFCHSSHQHDGMEAYVYVDSCEPSPQLMKNNRBAEDYDDLTDSEMDVVRP	379
QY	361	DDNSPSFQIRSVAKKPKTWHVYITAAEEDWDYAPLVPDPRSKSYQLNNGPQRTG	420
DB	380	DDNSPSFQIRSVAKKPKTWHVYITAAEEDWDYAPLVPDPRSKSYQLNNGPQRTG	439
QY	421	RKYKVRFWAYTDETPKTRERAOHESGILGLLYGSGVDTLLIIFKNQASRPNIYPHGI	480
DB	440	RKYKVRFWAYTDETPKTRERAOHESGILGLLYGSGVDTLLIIFKNQASRPNIYPHGI	499
QY	481	TDVRPLYSRRLPKGVKHLKDFPILPGEIFPKYKWTVTVEDGPTKSDPRCLTRYSSFWNME	540
DB	500	TDVRPLYSRRLPKGVKHLKDFPILPGEIFPKYKWTVTVEDGPTKSDPRCLTRYSSFWNME	559
QY	541	RLASGLIGPLLCYKESVDQRGNQIMSKBNVILSVFDEKESWYLTENIORFLPNPAG	600
DB	560	RLASGLIGPLLCYKESVDQRGNQIMSKBNVILSVFDEKESWYLTENIORFLPNPAG	619
QY	601	VOLEDPEFOASNIMHISINGVDFSLQLSVCLHEVAYWYILSIGAQDTDFLSVFFSGYTFKH	660
DB	620	VOLEDPEFOASNIMHISINGVDFSLQLSVCLHEVAYWYILSIGAQDTDFLSVFFSGYTFKH	679
QY	661	KMYVEDTLTLPFPGSGEIVFMSMENPGLWILGCHNSDFRNRGMTALLKVSQCDKNTGDYVE	720
DB	680	KMYVEDTLTLPFPGSGEIVFMSMENPGLWILGCHNSDFRNRGMTALLKVSQCDKNTGDYVE	739
QY	721	DSYVEDISAVLLSKNNAIEPRSP-----	742
DB	740	DSYVEDISAVLLSKNNAIEPRSPSTRQKQFNATTIPENDIEKTDPMFAHRTMP	799
QY	743	-----	742
DB	800	KIQNVSSDLLMLLRQSPHGLSLDLQEKYETSDDDPSGCAIDSNNSLSEWTHFRPQ	859
QY	743	-----	742

```
Db 860 LHSGDMVFTPEGLQLRLNEKLGTTAATLKLDFKVSSTSNLSTIPSDNLAAGTDN 919
Qy 743 ----- 742
Db 920 TSSLGPPSPMPVHYDSQDSTTLFGKKSPLTESGGPLSLSEBNDKSLLESGLMNSQSSW 979
Qy 743 ----- 742
Db 980 GKVSSTESGRLFKGRAHGPALLTKDNALFKVSIILLTKNTKSNNSATNRKTHIDGPSL 1039
Qy 743 ----- 742
Db 1040 LIENSPSYQWILESDTEFFKVTPLIHDRMLMDKNATRLNHNKNTTSSKNHMQVK 1099
Qy 743 ----- 742
Db 1100 KEGIPDDAQNPMSPFKMLFLPESARWIQRTHGKNSLNSGQSPKQLVSLGPEKSVEG 1159
Qy 743 ----- 742
Db 1160 QNFLSEKXKVVVGKEFTKDVGLKEMVFPSSRNLFITLNDLHNHNTNQBKKIQEIEK 1219
Qy 743 ----- 742
Db 1220 KETLIQENVVLPIHTVTGTGNFMKNLFLSTRQNVESYDGAAYAPVLQDFRSLNDSTNR 1279
Qy 743 ----- 742
Db 1280 TKKHTAHFSKGBENLEGLNQTKQIVKQVACTTRISPNTRSQQNFVTVQSRKALKQPL 1339
Qy 743 ----- 742
Db 1340 PLEETELEKRIIVDDTSTQSKNMKHLTPSTLTQIDYNEKEKGAITQSPSLDCLTRSHSI 1399
Qy 743 ----- 742
Db 1400 PQANRSPLIATKVSFPISIRPIYLRVLFDQNSHLPAASYRKDQSGVQESSHFLQCAK 1459
Qy 743 ----- 742
Db 1460 NNLSLAILTLEMTQDREVGLSGTSATNSVYKYKVENTVLPKPLPKTSKGVELLPKVI 1519
Qy 743 ----- 742
Db 1520 YQDLFPTETSGSPCHLDLVEGSLLOQTEGAIKWNEANRPGKVPFLRVATESAKTPSK 1579
Qy 743 ----- 742
Db 1580 LLDPLAMDNHYGTQIPKEBWSQBKSPEKTAFFKKKDTTILSNACESHAAJAINBQKNP 1639
Qy 743 -----SONPVLKXHQREITRTTLQSDQEEIDYDDTISVEMKKEDFDIY 786
Db 1640 EIEVTWAKQGRTERLCSQNPVLRKHQREITRTTLQSDQEEIDYDDTISVEMKKEDFDIY 1699
Qy 787 DEDENQSPRSFOKTRHYFIAAVERLDWYGNSSSPHVLNRNAQSGSVQFKKKVQFQETD 846
Db 1700 DEDENQSPRSFOKTRHYFIAAVERLDWYGNSSSPHVLNRNAQSGSVQFKKKVQFQETD 1759
Qy 847 GSFTQPLYRGLNEHLGLLGPYIRAEVDNIMVTFRNQASRPYSYSLISYEEDQORQA 906
Db 1760 GSFTQPLYRGLNEHLGLLGPYIRAEVDNIMVTFRNQASRPYSYSLISYEEDQORQA 1819
Qy 907 EPRKNFVKNETKTYFWKQVHMAPTKQDFCKAWAYFSDVDLEKDVHSLGILLVCHT 966
Db 1820 EPRKNFVKNETKTYFWKQVHMAPTKQDFCKAWAYFSDVDLEKDVHSLGILLVCHT 1879
Qy 967 NTLNPAHGRQVTVQBFALFFTFIDETKSWYFTFENMERNCRAPCNIQMEDPTFKENYRHA 1026
Db 1880 NTLNPAHGRQVTVQBFALFFTFIDETKSWYFTFENMERNCRAPCNIQMEDPTFKENYRHA 1939
Qy 1027 INGYIMDTLPGLVMAQDQIRWYLLSMGSNENIHSIHFSGHVFTVRKKEEYKQALYNLYP 1086
Db 1940 INGYIMDTLPGLVMAQDQIRWYLLSMGSNENIHSIHFSGHVFTVRKKEEYKQALYNLYP 1999
```

```
Qy 1087 GVFETVEMLPKAGIMRVECLIGEHLAGMSTLFLVYSNKCQTPLGMSGHIRDFOITAS 1146
Db 2000 GVFETVEMLPKAGIMRVECLIGEHLAGMSTLFLVYSNKCQTPLGMSGHIRDFOITAS 2059
Qy 1147 GQYQWAPKARLHYSGSINAWSTKEPFSWKVDLLAPMIHIGIKTOGARQKFSLSYISQ 1206
Db 2060 GQYQWAPKARLHYSGSINAWSTKEPFSWKVDLLAPMIHIGIKTOGARQKFSLSYISQ 2119
Qy 1207 FIIMYSLDGKKWQYRGSTGLMVFNGVNDSSGKINIFNPPIIARIYRILRHPHYISRS 1266
Db 2120 FIIMYSLDGKKWQYRGSTGLMVFNGVNDSSGKINIFNPPIIARIYRILRHPHYISRS 2179
Qy 1267 TLRMELMGCDLNSCMPLGMSKASISDAQITASSYFTNMFATNFSKARLHLQGRSNAWR 1326
Db 2180 TLRMELMGCDLNSCMPLGMSKASISDAQITASSYFTNMFATNFSKARLHLQGRSNAWR 2239
Qy 1327 PQVNNPKEWLQVDFQKTMKVTVTTQGVKSLTSMYVKEFLISSQGHQHWTLFFQNGKV 1386
Db 2240 PQVNNPKEWLQVDFQKTMKVTVTTQGVKSLTSMYVKEFLISSQGHQHWTLFFQNGKV 2299
Qy 1387 KVFGQNDSTPPVNSLDPLLTLYLRIHPQSWHQAIALRMEVLGCEAODLY 1438
Db 2300 KVFGQNDSTPPVNSLDPLLTLYLRIHPQSWHQAIALRMEVLGCEAODLY 2351
```

## RESULT 2

T42763

coagulation factor VIII precursor - pig

C:Species: Sus scrofa domestica (domestic pig)

C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 09-Jun-2000

C:Accession: T42763

R:Lollar, P.

submitted to the EMBL Data Library, August 1996

A:Reference number: Z22269

A:Accession: T42763

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: mRNA

A:Residues: 1-2133 &lt;COL&gt;

A:Cross-references: EMBL:U49517; NID:g1511633; PID:g1511634; PIDN:AB06705.1

C:Superfamily: coagulation factor VIII; discoidein 1 amino-terminal homology; ferroxi-

C:Keywords: acute phase; blood coagulation; duplication; glycoprotein; hemophilia A; i-

F:1-19/Domain: signal sequence #status Predicted &lt;SIG&gt;

F:20-2133/Product: coagulation factor VIII #status predicted &lt;MAT&gt;

F:223-349/Domain: ferroxidase repeat homology &lt;FOX1&gt;

F:402-730/Domain: ferroxidase repeat homology &lt;FOX2&gt;

F:1498-1820/Domain: ferroxidase repeat homology &lt;FOX3&gt;

## Query Match

Best Local Similarity 81.0%; Score 6227; DB 2; Length 2133;

Matches 1212; Conservative 106; Mismatches 119; Indels 678; Gaps 3;

```
Qy 1 ATRYYILGAVELSDWYMQSD-LGELPYDARPPRPVPSFPNTSVVYKTLFVETVHLF 59
Db 20 AIRYYILGAVELSDWYRQSELLRELVDTTPATAPGALPLGPSLVYKTVFVEFTDQLF 79
Qy 60 NIAKPRPMWLLGPTTQAEVYDVTVTILKNMASHPSVLSHAGVSYWKASGEYDDOTS 119
Db 80 SVAREPPPMWLLGPTTQAEVYDVTVTILKNMASHPSVLSHAGVSYFWKSSGEYEDHTS 139
Qy 120 QREKDDKVPFGSGHYYVWVLKENGPMASDPLCLITYLSYLSHVDLVKDLNGLIGALLVC 179
Db 140 QREKDDKVLPGKSTQVWVLKENGHTASDPLCLITYLSYLSHVDLVKDLNGLIGALLVC 199
Qy 180 REGSLAKEKTOQLHKFILLPAVFDEGKSWHSETKNSLMQDRDAASARAWPKHTVNGYVN 239
Db 200 REGSLTRERTQNLHEFVLLPAVFDEGKSWHSARNDSTWTRAMDPAAPAQPAHMTVNGYVN 259
Qy 240 RSLPLGLICHRKSKSYWVHGMGTTPVHVSIFLEGHTFLVRNHRQASLEISIFLTAQTL 299
Db 260 RSLPLGLICHRKSKSYWVHGMGTSPVHVSIFLEGHTFLVRNHRQASLEISIFLTAQTF 319
Qy 300 LMDLGGQFLLFCHISHQHDGMEAYVKVDSCEPEEQRLRMKNNEAEDYDDDLTDSEMDVVR 359
```





Query Match	80.5%;	Score 6192;	DB 2;	Length 2319;	
Best Local Similarity	53.4%;	Pred. No. 0;			
Matches 1231;	Conservative	97;	Mismatches 103;	Indels 876;	Gaps 9;
Qy	1	ATRYVILGAVELSWDYMOSD- LGE L PVDARPPPRVPKSPFNTSVVYKTLFVFEFTWHLF	59		
Db	20	AIRRYVILGAVELSWNVIQSDLLSVLHTDSRFLPRMSTSPFNTSIMYKTVFVYKQDLF	79		
Qy	60	NIAPRPPMGLLGPTTQAEVDVTWITLKNMASHPVSLHAGVSVYWKASGEGABYDDOTS	119		
Db	80	NIAPRPPMGLLGPTTWTWHTVITLKNMASHPVSLHAGVSVYWKASGEGABYDDOTS	139		
Qy	120	QREKEDKVPFGGSHYTVVMOVLKENGPMASDPCLTYSLSHVDLVKDLNSGLIGALLVC	179		
Db	140	QMEKEDKVPFGESHYTVVMOVLKENGPMASDPCLTYSYMSHVDLVKDLNSGLIGALLVC	199		
Qy	180	REGSLAKETQTLHKFTILLFAVDEGKSWHSETKNSLMQORDAASAPAMGHTVNGYVN	239		
Db	200	KEGSLSKERTQMLYQFVLLFAVDEGKSWHSETNDSYTSQMSDASARDWPMHTVNGYVN	259		
Qy	240	RSPLGLIGCHRKSVYHVIQMGTTPEVHSIFLEHTEFLVRNHRQASLEISPIFLTAQTL	299		
Db	260	RSPLGLIGCHRKSVYHVIQMGTTPEIHSIFLEHTEFLVRNHRQASLEISPIFLTAQTL	319		
Qy	300	LMIDGQFLLFCHISSHQHDMGEAYVKVDSCPEEPQLEMK- NNEEAEDYDDDLTDSMDVV	358		
Db	320	LIDLGQFLLFCHISSHKHDMGEAYVKVDSCPESSQWKNNNEEWEYDDDLV- SEMDMF	378		
Qy	359	RFDDNSPFIQIRSAVAKGPKTWVHYIAABEEDWDYAPLVAPDDRSYKSOYLNNQPR	418		
Db	379	TLDYDSSP- FIQIRSAVAKPKTWIHYISABEEDWDYAPSVPTSDNGSYKSOVLSNGPFR	437		
Qy	419	IGRKYKVRFWAYDETFTKREAOHESGILGPLLYGEVGDTLIIIPKNOASRPYNIYPH	478		
Db	438	IGRKYKVRFWAYDETFTKRETIQHESGLGPLLYGEVGDTLIIIPKNOASRPYNIYPH	497		
Qy	479	GITDVRPLYSRRLPKGVKHLKDFPILPCEIFKYKWTVTVEDGPTKSDPRCLTRYSSFN	538		
Db	498	GITDVSPLHARLPGRGIKHVKOLPIHPGEIFKYKWTVTVEDGPTKSDPRCLTRYSSFN	557		
Qy	539	MERDLASGLIGPLLI CYKESVDQGNQIMSDKRNVLIFSPVDENRWSYLTENQRFPLNP	598		
Db	558	PERDLASGLIGPLLI CYKESVDQGNQIMSDKRNVLIFSPVDENRWSYLTENQRFPLNP	617		
Qy	599	AGVQLEDPBEOASIMESINGYVDSLSVCLHEVAYWYTLTGATDPLSVFSGVTP	658		
Db	618	AKTQPDGFGQASIMESINGYVDSLSVCLHEVAYWYTLTGATDPLSVFSGVTP	677		
Qy	659	KHKMAYEDTLTLPFSGETVPMSENPCMLTILGCHNSDFRNRMGTALLKVSSCDKNTGDY	718		
Db	678	KHKMAYEDTLTLPFSGETVPMSENPCMLTILGCHNSDFRNRMGTALLKVSSCDKNTSDY	737		
Qy	719	YEDSYEDISAVLLSKNNAIEPRFSQON	745		
Db	738	YEETIYEDIPTQLVNNENNVIDPRSPFQNTNHPNTRKKFKDSTIPKNDMEKIEPOFEBIAE	797		
Qy	746		745		
Db	798	MLKVQSVSVSDMLLQGSHPTPHGLFLSDGQEAIEYRAIHDDHSPNAIDSNEGSPKVTQL	857		
Qy	746		745		
Db	858	RPESHSEKIVFTPOPGLRSNKSLETTIEVKWKKLGLQVSSLPNSLMTTILSDNLKA	917		
Qy	746		745		
Db	918	TFEKTDSGFPDMPVHSSSKLSTTAGKKAYSLVGSHPVPLNASEBNSDNLDTSLMYSQ	977		
Qy	746		745		
Db	978	ESLPRDNLSTENDRLREKRFHGIALTKONTLTKONVSLMKTNTYHSTTNEKLHTE	1037		
Qy	746		745		
Db	1038	SPTSIENTSTDLQDAILKVNSEIQEVTALIHDTGLGKNSYTLRLNMLNRTTSTKNKDI	1097		
Qy	746		745		
Db	1098	FHRKDEDDPIQDEBENTIMPFKMLFLSESSNFKKTNGNNSLNSQEHSPKQVLVLMFKK	1157		
Qy	746		745		
Db	1158	YVKNSQFLSEKKNVTVEQDFTKNIGLKDMPAPPHNMSIFLTTLNSVHENGHRHNOEKNIQE	1217		
Qy	746		745		
Db	1218	EIEKEALIBEKVVLPQVHEATGSKNFLKDILILGTRQNISLYEHVHPVLQNIITSINNSTN	1277		
Qy	746		745		
Db	1278	TVOIHMEHFFKRKDKETNSBGLVNTREKVKYPSQKNITTORSKRALQOFLRSTQWLK	1337		
Qy	746		745		
Db	1338	TINCSQCIILKQIDHSKEMKCKPITKSSLSLSDSSVVKSTQTNSSDSHIVKTSAPPPIDLK	1397		
Qy	746		745		
Db	1398	SPFQNKFSHVQASSYIYDFKTKSSRIQESNNFLKETKINNPSLAILPWNMFIDQKFTSP	1457		
Qy	746		745		
Db	1458	GKSNSTSVTYKKRENIIFLKPITLPSBSGKIELLPQVSIQEEIILPTSTSHSGPGLNLMK	1517		
Qy	746		745		
Db	1518	EVFLQIKQPTKWKAKRHGESIKGTESSKNTSKLLNHHAWDYHAAQIPKDMWKSKE	1577		
Qy	746		751		
Db	1578	KSPETIISKQEDTILSLRPHGNSHISGANQWNPQRETTWVKQGTQRTCSQITPPVLKR	1637		
Qy	752	HQREITRTTLTQSDQEEIDYDITISVEMKEDPDIDYDENQSPRSFOKTRHYFIAVER	811		
Db	1638	HQREL--SAPQSOEATDYDAILTIE--TIEDFDIYSEDIKQSPRSFOKTRHYFIAVER	1694		
Qy	812	LWDYGMSSSHVLRNRAQSGSVQFKKVPQEFQSDGFTQPLYRGELNEHLGLGPYIRA	871		
Db	1695	LWDYGMSTS--HVLNRVQSDNVQFKKVPQEFQSDGFTQPLYRGELNEHLGLGPYIRA	1753		
Qy	872	EVEDNIMVTFRNQAASRPSYSSLSIYEEEDQOQAEPRKPVKPNETKTYFWKVQHMAP	931		
Db	1754	EVEDNIMVTFRNQAASRPSYSSLSIYKEDQR--GEPRNRFVKPNETKIYFWKVQHMAP	1812		
Qy	932	TKDEPDCWAYPSDVLKDVHSGSLGPLLVCHTNTLNPAGHGVTVQSFALFTTFDE	991		
Db	1813	TEDEFDCWAYPSDVLKDVHSGSLGPLLVCHTNTLNPAGHGVTVQSFALFTTFDE	1872		
Qy	992	TKSWYFTENNERCNPACNTQMEDPTPKENYRFAHNGYIMDTPLGLVMAQDQIRWYLL	1051		
Db	1873	TKSWYFTENNERCNPACNTQMEDPTPKENYRFAHNGYIMDTPLGLVMAQDQIRWYLL	1932		
Qy	1052	SMGNSNENIHSFSGHYFTVRKKEEYKALYNLYPGVFETVEMLPKAGIWRVECLIGEH	1111		
Db	1933	SMGNNENIQSIHFSGHVFTVRKKEEYKALYNLYPGVFETVEMLPKAGIWRVECLIGEH	1992		
Qy	1112	LHAGMSTFLVYKNCQOTPLGMASGHTRDQITASQYQOWAPKLARLHYSGSINAWSTK	1171		
Db	1993	LQAGMSTFLVYKNCQOTPLGMASGHTRDQITASQYQOWAPKLARLHYSGSINAWSTK	2052		
Qy	1172	EPSPWIKVDLLAPMI IHGKITQAGARQKFSLSYISQFIIMYSLDGKKTWOTYRGSTGTLMV	1231		
Db	2053	EPSPWIKVDLLAPMI IHGKITQAGARQKFSLSYISQFIIMYSLDGKKTWOTYRGSTGTLMV	2112		
Qy	1232	FFGNVDSSGKHNIFNPPIIARIHLPHYTHYSIRSTLRMBELMGCDLNSCMPLGWESKAI	1291		

```

Db      2113  PFGVDSGKXNSFNPIIARIYRLHPTHTSSIRSTRMLWGLCDLNSCSIPLGWESKVI 2172
Qy      1292  SDAQITASSYFTNMPATWSPKARHLQGRSHNAPQVNNPKENIQLVDFTKMTKVTGTT 1351
Db      2173  SDTQITASSYFTNMPATWSPKARHLQGRSHNAPQVNNPKENIQLVDFTKMTKVTGTT 2232
Qy      1352  QGVKSLTSMVKEFLISSQDGHOWTLFFQNGKVKVFGQNDSTFPVNSLDPPLLTTRY 1411
Db      2233  QGVKSLTSMVKEFLISSQDGHOWTLFFQNGKVKVFGQNDSTFPVNSLDPPLLTTRY 2292
Qy      1412  LRIHPQSWHQIALRMEVLGCEAQLY 1438
Db      2293  LRIHPQIWEHQIALRMEVLGCEAQLY 2319

RESULT 4
KFHUS
N:Alternate names: coagulation factor V precursor [validated] - human
C:Species: Homo sapiens (man)
C:Date: 19-May-1989 #sequence revision 02-Jun-1995 #text_change 08-Dec-2000
C:Accession: A56172; A42344; A28028; A27498; A25897
R:Cripe, L.D.; Moore, K.D.; Kane, W.H.
Biochemistry 31, 3777-3785, 1992
A:Title: Structure of the gene for human coagulation factor V.
A:Reference number: A42344; MUID:92232668; PMID:1567832
A:Accession: A56172
A:Molecule type: DNA
A:Residues: 1-2224 <C>
A:Cross-references: GB:J05368
A:Accession: A42344
A:Molecule type: DNA
A:Residues: 48-58; 79-89; 120-130; 191-201; 239-249; 313-323; 368-378; 428-437; 461-471; 533-542;
2070; 2111-2120; 2172-2181 <C2>
R:Jenny, R.J.; Pittman, D.D.; Toole, J.J.; Kriz, J.W.; Aldape, R.A.; Hewick, R.M.; Kaufm
Proc. Natl. Acad. Sci. U.S.A. 84, 4846-4850, 1987
A:Title: Complete cDNA and derived amino acid sequence of human factor V.
A:Reference number: A28028; MUID:87260886; PMID:3110773
A:Accession: A28028
A:Molecule type: mRNA
A:Residues: 1-857; 'R', 859-864, 'R', 866-924, 'E', 926-1763, 'I', 1765-2212, 'T', 2214-2224 <JEN>
A:Cross-references: GB:M16967
A:Note: Parts of this sequence, including the amino end of the mature protein, were dete
R:Kane, W.H.; Ichinose, A.; Hagen, F.S.; Davie, E.W.
Biochemistry 26, 6508-6514, 1987
A:Title: Cloning of cDNAs coding for the heavy chain region and connecting region of hum
A:Reference number: A27498; MUID:88107560; PMID:2827731
A:Accession: A27498
A:Molecule type: mRNA
A:Residues: 1-1284, 'I', 1286-1600 <KAN>
A:Cross-references: GB:M17785
A:Note: Parts of this sequence were determined by protein sequencing
R:Kane, W.H.; Davie, E.W.
Proc. Natl. Acad. Sci. U.S.A. 83, 6800-6804, 1986
A:Title: Cloning of a cDNA coding for human factor V, a blood coagulation factor homolog
A:Reference number: A25897; MUID:86313665; PMID:3092220
A:Accession: A25897
A:Molecule type: mRNA
A:Residues: 1188-1215, 1315-2224 <KA2>
A:Cross-references: GB:M14335
A:Note: Parts of this sequence were determined by protein sequencing
R:Keller, F.G.; Ortal, T.L.; Quinn-Allen, M.A.; Kane, W.H.
Biochemistry 34, 4118-4124, 1995
A:Title: Thrombin-activated activation of recombinant human factor V.
A:Reference number: A56139; MUID:92210278; PMID:7696276
A:Contents: annotation; thrombin cleavage sites
C:Comment: Factor V is activated by thrombin and partially by coagulation factor Xa.
C:Genetics:
A:Gene: GDB:F5
A:Cross-references: GDB:1119896; OMIM:227400
A:Map position: 1q23-1q23
A:Introns: 53/2; 84/1; 125/1; 196/1; 244/1; 318/1; 373/2; 432/3; 466/1; 537/3; 588/1; 65

```

```

C:Function:
A:Description: acts as a cofactor, with calcium and phospholipid, for the factor Xa p
A:Pathway: blood coagulation
C:Superfamily: coagulation factor V; discoidein I amino-terminal homology; ferroxidase
C:Keywords: blood coagulation; duplication; glycoprotein; phospholipid binding; plasme
F:1-28/Domain: signal sequence #status predicted <SIG>
F:19-2224/Product: coagulation factor V #status predicted <MAT>
F:29-737/Product: coagulation factor V heavy chain #status experimental <VAH>
F:29-345/Domain: A1 <DA1>
F:33-329/Domain: ferroxidase repeat homology <FO1>
F:346-691/Domain: A2 <DA2>
F:351-684/Domain: ferroxidase repeat homology <FO2>
F:692-1573/Domain: B <DOB>
F:1183-1461/Region: 9-residue repeats (Q-X-T/N-L-S-P-D-L-S)
F:1574-2224/Product: coagulation factor Va light chain #status experimental <VAL>
F:1574-1905/Domain: A3 <DA3>
F:1581-1905/Domain: ferroxidase repeat homology <FO3>
F:1667-1765/Region: phospholipid binding #status predicted
F:1906-2064/Domain: C1 <DC1>
F:2065-2224/Domain: C2 <DC2>
F:2065-2221/Domain: discoidein I amino-terminal homology <DN2>
F:51-55,239,297,460,468,554,741,752,760,776,782,821,938,977,1074,1083,1103,1106,1479,
F:167-193,248-329,500-526,603-684,1725-1751,1907-2061,2066-2221/Dissulfide Bonds: #sta
F:334-335/Cleavage site: Arg-Asn (protein C) #status predicted
F:363,693,1546/Binding site: sulfate (Tyr) (covalent) #status predicted
F:376-377/Cleavage site: Arg-Ser (coagulation factor Xa) #status predicted
F:382,1338/Binding site: carboxylate (Asn) (covalent) #status absent
F:534-535/Cleavage site: Arg-Gly (protein C) #status predicted
F:737-738/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
F:1046-1047/Cleavage site: Arg-Thr (coagulation factor Xa, thrombin) #status experime
F:1573-1574/Cleavage site: Arg-Ser (thrombin) #status experimental

Query Match      31.1%; Score 2390.5; DB 1; Length 2224;
Best Local Similarity 26.1%; Pred. No. 6.4e-151;
Matches 593; Conservative 277; Mismatches 483; Indels 923; Gaps 35;

Qy      3  RRYLGAVELGWDVYMQSDGLGELPVDARPPRPVPSFPENTSV-YKTLFVFTVHLFNI 61
Db      32  RQFYVAAQGISWSRPE-----PTNSLNLSVTSFKKIVREYEPY-PKK 75

Qy      62  AKRPPWMLGTLGPTQARVYDTVTITLKNMASHVSLHAGVYVYKASGEGADYDQTSOR 121
Db      76  EKQSTISGLLPTLYAEVDGIIVKFKKADKPLSIHPQIRYSKLSGASLYLDTTPA 135

Qy      122  EKEDDKVPFGSHTVVOVLKENGPMASDPCLTYSYLSHVDLVKDLNSGLIGALIVCE 181
Db      136  EKMDDAVAPGREYTYWESISDSGTHDDPPCLTHIYYSHENLIEDFNSGLIGLILCKK 195

Qy      182  GSLAKEKTQ-TLHK-FILLFAVFDGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYV 239
Db      196  GLTSGGTQKTFDQIVLLFAVFDGKSWSQSS-----LMTVNGYV 239

Qy      240  RSLPLGIGCHRSVYVHVGMGTPPEVHSIFLEGHTFLVRNHRQASLEISPTFTTAQTL 299
Db      240  GTMPDITVCAHDHISWLLGSSGPELFSIHFGQVLEQNHKVSATILVSATSTANMT 299

Qy      300  LMDLQFLLFCHISSHQHDGMEAVYKVDSCPEEPQLRMKNNEAEYDLDLTDSEMDVR 359
Db      300  VGFEKWIISITTPKHLQAGWQAYIDKNCPKKTKNLKKTRE----- 342

Qy      360  FDDNDSPSFIQIRSVAKKPKTWVHYIAAEEDWDYAPLVLAPDDRSYKSOYINNGPQRI 419
Db      343  -----QRREKWEVFIAAEEVWDYAPVIANMDKYSRSHLDNFSNQI 387

Qy      420  GRKYKVRFMAYDTEFTKTRAI-----QHEGILGPLLYGVEGDTLLIIPKQASRPYNI 476
Db      388  GRHYKVMYQYEDDSF-TKHTVNPMMKEDGILGPIIRAQVRDTLKIIVFKNMASRPYSI 446

Qy      477  PHGIT-----DVRFLYSRLPKGVKHLKDFILPGEIFKYKWTIVVEDGPTKSPRCIT 530
Db      447  PHGVTFSPYEDVNGSFT-----SGRNTMTIRAVQGETYTYKWNILEDFDEPTENDACQCL 502

```

QY 531 RYSSFNVERDASGLGPIILCYKESVDQGNQIMSDKXNVILFSPVDENRSLTEN 590  
Db 503 RPYSDVIMDIDIASGLIGLLICKSRSLDRGIQRAADI EQAVFAVDFENKSWYLEDN 562  
QY 591 IQREFLPAGVQLEDPEPEQASINMSINGYVFDLSL-QLSVCLHEVAYVYVILSIGAQTDPL 649  
Db 563 INKFCENDEVKRDDPKYESNIMSTINGYVPESTITLGFCEDDIVQHFCSVGTQNEIL 622  
QY 650 SVFSGYTFKHMVYEDTLTLPFSGEYVFMENPGLWILGCHNS----- 695  
Db 623 TIHPTGHSFYKGRHEDTLTLPFMRGESVTVMNDVGTWMLTSMNSSPRSKKLRLKPRDV 682  
QY 696 -----DPENR----- 700  
Db 683 KCIPODDSDSYEIFEPPSTVMATKMDRLPEDEDESADYDQNRLLAALGIRSPRNS 742  
QY 701 -----GMTAL----- 705  
Db 743 SINQEEFENLTALENGTEFVSNTDII VGSNYSSPSNISKFTVNNLAEPOKAPSHQ 802  
QY 706 -----LKVSSCDKNTGYDESDYED----- 725  
Db 803 ATTAGSPHLHIGKNSVLNSTAEHS SPYSDPIEDPLQPDVGTGIRLLSLGAGBPKSQEH 862  
QY 726 ----- 725  
Db 863 AXHKGPKVERDQAACHRFSSWMLLAHKVGRHLSQDTSPSGMRPWEDLFSQDTSPSRMR 922  
QY 726 -----ISAYLLSKNA----- 736  
Db 923 PKWDPSPDLLLLKQSNSSKILVGRWHLASEKGSVEIIQDTEDTAVNNWLLISPOQASRAW 982  
QY 737 ----- 736  
Db 983 GESTPLANKPGQSGHPKPRVRHKSQVRDGGKSLKKSQFLIKTRKKKKKXHTHAP 1042  
QY 737 IEPSPF-----SQ 744  
Db 1043 LSPRTFHLRSEAYNTFSERLLKHSLVHLKSNETS LPTDLNQTLPFMDFGHIASLPDHNQ 1102  
QY 745 N----- 745  
Db 1103 NSSNDTQASCPGLGYQTVPEEHYQTFPIQDPDQMHSTSPSHRSSSPSELMLEYDRS 1162  
QY 746 -----PVLKXH- 752  
Db 1163 HKSPTDISQSPSSSEHEVQTVISPDLSQVTLSPELSQTNLSPDLSHTTILSPELIQRL 1222  
QY 753 ----- 752  
Db 1223 SPALGQMPIPDLSHTTILSPDLSHTTILSLDLSQTNLSPELSQTNLSPALGQMPIPDLSH 1282  
QY 753 ----- 752  
Db 1283 TTLSLDFSQTNLSPELSHMTLSPELSTQNLSPALGQMPIPDLSHTTILSLDFSQTNLSP 1342  
QY 753 -----QREIRTKTILQSDQEBI----- 768  
Db 1343 LSQTNLSPALGQMPIPDPSHTTILSLDLSQTNLSPELSQTNLSPDLSHTTILSPADLSQIPL 1402  
QY 769 ----- 768  
Db 1403 TPDLDQMTLSPDLGETDLSPNFGQMSLSPDLSQVTLSPDISDTTLLPDLSQISPPDLQ 1462  
QY 769 -----DYDDT 773  
Db 1463 IFYPSESQSLLLOFENESFPYDPLGQMPSPSPPLNDTFLSKFNPLIVGLSKDQGYD 1522  
QY 774 ISVEMKKE-----DFDIYDE-----DENQSP-----RSFOKKTRHY 804  
Db 1523 IEIIPKEVQSSDEDDYABIDIVPDYDDPKYDVTWNSSRDPDNTAAWYLSNNGRRNY 1582  
QY 805 FIAAVERLWDYGMSSPHVLRNR--AQSGSVFQ----FKKVVVFQEFDTGDSFTQPLYRGLN 859

Db 1583 YIAAEISWDY-----SEFVQRETIEDSDDIPEDTTTKYKVVFRKYLSTFTKRDPRGRYE 1638  
QY 860 EHLGLGPIYIRAEVEDNIMVFRNQASRPYGFYSLSISYE-----EDORQCAEPRKN 911  
Db 1639 EHLGLGPIYIRAEVDDVIQVFRKNLASRPYSYLAHGLSYEKSSBGTVDDEDDSPFWFKEDN 1698  
QY 912 FVKQNETKTYFWKQVQHMAPTKDBFCDKAMAYFSDVLEKDVHSLGILGLPLLVCHTNTNLP 971  
Db 1699 AVQPNSSYTYVWHATERSGPSPGACRAMAYYSAVNPEDKIHGSLIGLPLLCQKGLHK 1758  
QY 972 AHGQVTVQEFALFTTIFDETKSNYFTFENMERNCRAPCNIQWEDPTFKENTRFHAINGYI 1031  
Db 1759 DSNPLDMREFVLLFMFTFDEKKSYWEKKSSSWR-----LTSEMKKSHEFHAINGM 1812  
QY 1032 MDTLPLGLVMAQDQIRWYLLSMGSENENIHS IHFSGHVFTVRKKEKYSKMALNLYPGVPE 1091  
Db 1813 Y-SLPLGKMYEQEWRVRLHLLNIGSGQDIHVHVFHGTLLNENKQHQGLGWPLLPGSFKT 1871  
QY 1092 VEMLPKAGIWRVRECLIGEHLHAGMSTLFLVYSNKCQTPPLGMASSHIRDFQITASGOYQG 1151  
Db 1872 LEMKASKPGWMLLNTVEGENQORAGWQTPFLIMDRDCRMPMGLSTGIISDSQIKASEFLGY 1931  
QY 1152 WAPKLARLHYSGSINASTKE---PFS---WIKVDLAPMIIHGIKTCGAKROKFSLLYIS 1205  
Db 1932 WEPLARLNNGSGYNASWSEKLAAPFASKPMIOWDMQKEVIITGIOTQGAHYLKSCTYT 1991  
QY 1206 QFIIMYLDGKKWQTYRGNSGTLMVFFGNVDSGIGHNIPNPPIIARYIRLHPHYR 1265  
Db 1992 EFYVAYSNQINWQIFKGNSTRNVMYFNGSDATIKENQDPPIVARIIRISPRAYNR 2051  
QY 1266 STLRLMELGCDLNSCMPLGMSKAIISDAQITASSYFTNMFEAT-WSPSKARLHLQGRNA 1324  
Db 2052 PTLRLLEQCEVNGCSTPLGMENKIKENQITASSFKKSWMGDYWEPRARLNAQGRVA 2111  
QY 1325 WRPOVNPKEWLQVDFQKTMKVTVGTTCGVKSLTSTMYRFLISSSDQGHQWTLFPQNG 1384  
Db 2112 WQAKANNKOWLEIDLKIKKIKITAITITQGCCKSLSEMYVYSYTIHYSQGVKPYRLKS 2171  
QY 1385 KV--KVFQGNQDSFTPVVNSLDPPLLTLYLRIHPQSWHQAIALRMEVLGCEAQDLY 1438  
Db 2172 SWDKIFEGNTYTKGHVKNFNPFIISRFIKVIPKTNQSLAKLLELFGC---DIY 2224

## RESULT 5

## KFBO5

coagulation factor V precursor - bovine

C/Species: Bos primigenius taurus (cattle)

C/Date: 04-Mar-1993 #sequence\_revision 28-Apr-1995 #text\_change 11-Jun-1999

C/Accession: A42580; A36497

R/Guinto, E.R.; Eamon, C.T.; Mann, K.G.; MacGillivray, R.T.

J. Biol. Chem. 267, 2971-2978, 1992

A/Title: The complete cDNA sequence of bovine coagulation factor V.

A/Reference number: A42580; MUID:92147638; PMID:11737753

A/Accession: A42580

A/Molecule type: mRNA

A/Residues: 1-2211 &lt;GUI&gt;

A/Cross-references: GB:W81440; NID:G163037; PIDN:AAA30512.1; PID:G163038

A/Note: sequence extracted from NCBI backbone (NCBIN:80774, NCBI:80776)

R/Kalafatis, M.; Jenny, R.J.; Mann, K.G.

J. Biol. Chem. 265, 21580-21589, 1990

A/Title: Identification and characterization of a phospholipid-binding site of bovine

A/Reference number: A36497; MUID:91072354; PMID:2254316

A/Accession: A36497

A/Molecule type: protein

A/Residues: 1566-1570, 'X', 1572-1581, 'X', 1583-1594, 1673-1676, 'X', 1678-1679, 'X', 1681, 'X',

R/Xue, J.; Kalafatis, M.; Silveira, J.R.; Kung, C.; Mann, K.G.

Biochemistry 33, 13109-13116, 1994

A/Title: Determination of the disulfide bridges in factor Va heavy chain.

A/Reference number: A55979; MUID:95034740; PMID:7947716

A/Contents: annotation

A/Note: 566-Cys and 617-Cys were shown to have free sulphydryls

C/Comment: Factor V is activated by thrombin and partially by coagulation factor Xa.

C;function:  
 A;Description: acts as a cofactor, with calcium and phospholipid, for the factor Xa prot  
 A;Pathway: blood coagulation  
 C;Superfamily: coagulation factor V; discoidin I amino-terminal homology; ferroxidase re  
 C;Keywords: blood coagulation; duplication; glycoprotein; phospholipid binding; plasma;  
 F;1-28/Domain: signal sequence #status predicted <SIG>  
 F;29-2211/Product: coagulation factor V #status predicted <MAT>  
 F;29-741/Product: coagulation factor Va heavy chain #status predicted <VAH>  
 F;29-345/Domain: A1 <DA1>  
 F;33-329/Domain: ferroxidase repeat homology <FO1>  
 F;346-695/Domain: A2 <DA2>  
 F;351-688/Domain: ferroxidase repeat homology <FO2>  
 F;696-1564/Domain: B <DOB>  
 F;1175-1437/Region: 9-residue repeats (Q-X-T-N-L-S-P-D-L-S)  
 F;1565-2211/Product: coagulation factor Va light chain #status predicted <VAL>  
 F;1565-1892/Domain: A3 <DA3>  
 F;1572-1892/Domain: ferroxidase repeat homology <FO3>  
 F;1654-1752/Region: phospholipid binding #status predicted  
 F;1893-2051/Domain: C1 <DC1>  
 F;1893-2048/Domain: discoidin I amino-terminal homology <DNI>  
 F;2052-2211/Domain: C2 <DC2>  
 F;2052-2208/Domain: discoidin I amino-terminal homology <DN2>  
 F;167-193,248-329,499-525/Disulfide bonds: #status experimental  
 F;225,239,297,382,460,553,587,745,756,774,780,902,952,964,1044,1053,1062,1071,1078,1094,  
 F;334-335/Cleavage site: Arg-Asn (protein C) #status predicted  
 F;363,697,1537/Binding site: sulfat (Tyr) (covalent) #status predicted  
 F;376-377/Cleavage site: Arg-Ser (coagulation factor Xa) #status predicted  
 F;533-534/Cleavage site: Arg-Gly (protein C) #status predicted  
 F;607-688,1712-1738,1894-2048,2053-2208/Disulfide bonds: #status predicted  
 F;741-742/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status predicted  
 F;1034-1035/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status predicted  
 F;1564-1565/Cleavage site: Arg-Ser (thrombin) #status experimental

Query Match 30.8%; Score 2372; DB 1; Length 2211;

Best Local Similarity 26.2%; Pred. No. 1.1e-149;

Matches 594; Conservative 265; Mismatches 487; Indels 924; Gaps 35;

QY	3	RYVILGAVLSVDYMQSDLGELPVDARFPPRPVKSPENTSVVYKTLFVETVHLENIA	62
Db	32	RFQVAAQIRNWR-----PESTHL-----SSPFETS--FKKIVRYEAY-FQKE	76
QY	63	KRPFPWMLGPTIOAEVYTVITLKNMASHPVSLHVGVSWKASGARYDDQTSORE	122
Db	77	KQSRSTGLGLPTLYAEVGDIMKVPKKAHPLSLHAQIKYKFSGASYSDDLPMWE	136
QY	123	KEDDKVPQGSHTYVWVULKEGPMASDPLCTYSVLSHVDLVKOLNSGLIGALLVCREG	182
Db	137	KMDDAVAPQGYTYEWTISEHSGPTHDDPPCLUTHIYYSYVNLVEDFNSGLIGELLICKKG	196
QY	183	SLAKEKTQTL--HKFTLLFAVDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNR	240
Db	197	TLTDEGTQKMPKQHVLMFAVDEKSNWQTS-----LMTVNGYNG	240
QY	241	SLPGLIGCHRKSVYWHVIGMGTTPPEVHSIFLEGHTFLVNRHQASLEISPIITFLTAQTLL	300
Db	241	TWPDITVCAHDHISWELIGMSSGPFLSFHFGQVLEQNHHKISAITLVSATSTTANMTV	300
QY	301	MDLGQFLPCHLSSHQHDGMEAYVKVDSCPEPQPMKONNEAEYDDDLTSEMDVVRP	360
Db	301	SPEGRWTIASLIPRHFGQAGQAYIDIKCAKTRNPK-----LTRDQ-----	343
QY	361	DDNDSFPTQIRSVAKKPKTVHYIAAEEEDWDVAPLVLPDDRYSKSOYLNNGPQIG	420
Db	344	-----RRHKRWBYFAAEVIMDYAPIIPANMDKKIRSLHLONFNRIG	388
QY	421	RYKVKVRMAYTDETKTR--BAIQHESGILGPLLYGEVGDITLLIFKNQASRPYNIYPH	478
Db	389	KHYKVVYKQYQDSTFKLEDPSSSEGDGILGPIIPAQVQVDTLKIVFKNWSRSYSIYPH	448
QY	479	GIT-----DVRPLYSRRLPKGVKHLKDPFLLPGEIFKFKYKWTVTVEGPKSDPRCL	529
Db	449	GVTFSYDNEVNSSTSGSNTMIAVR-----PGETYTYKWNILESDEPTENDAQCL	500
QY	530	TRYYSFVNMERDLASGLIGLIPLLICYKESVDQRGNQIMSDKKNVILFVSFVFNRRSWYLTE	589
Db	501	TFPYYSNDVITRDLAGLIGLILCKSRLSGRQIADIEQAVFAVFNKSWYIED	560
QY	590	NIQRPPLNPAGVQLEDDEPFOASNMH-----SINGYVFDLSQ-LSVCLHEVAYWYLSIG	643
Db	561	NIYKFCENPEKVRDDPKFYESNIMSNFTLPAINGYVPESIPILGFCFDDTVQMHFCSVG	620
QY	644	AQTDPLSVFSGYTFKKHMYVEDTLTLFPFSGEYVFMENPGLMILCHNSDFENRGMT	703
Db	621	TQNDLTLTHFCHSFTYGRKHEDTLTLFPMQGESVTVTMDNVGTWMLTMTNSNPKKL	680
QY	704	ALLKVSSCDKNTGD-----YYEDS-----YEDISAYLL-----	731
Db	681	LRFRDAKICIRDDDDSYELIYEPGSGTAMTYKKIHDSIEEDENDADSDYQDELALILGL	740
QY	732	-----SKNNAIEPRSFQ-----	744
Db	741	RSFRNSLNQEKDELNTALALEKDSFIPPSANRSLDSNSSRSHVSRLLIAKNAESILK	800
QY	745	-----NPPVLK-----	750
Db	801	TLHLLEAPAGSPLEHAGLDKNSALNPPMAEHSPPSYSEDPREDHPFLSDVTGVSLLPFGTG	860
QY	751	-----RHQR-----EITRTTLQSDQ-----EEI-----	768
Db	861	FKNRKPAKHQRFQVGRGOAAKHKFSQTRFPAKHKTTLRLSQDNSSSRMGWEDIFSDLLL	920
QY	769	-----	768
Db	921	LQKDPYKILNGEHLVSEKSGYIIOANENKNTVKNLPNSPQNDRTWGENIPFKNSHG	980
QY	769	-----	768
Db	981	KQSGHTFLVTRKPLQDRQRNRSLKEGLPLINTRKKKKEKPAYHVLSPRSFHLR	1040
QY	769	-----DYDD-----	772
Db	1041	GEVNASFDRRRHNSLLHASNETSLSIDLNQTFPSMNLSLAASLPDHDQTSNDTSTQT	1100
QY	773	-----TISVEMKKEDFDIYDE-----ENQSP-----	794
Db	1101	SSPDLPTVSPPEHYQIFPIQDSPTHTTAPSNRSPDPTHTTAPSNRSPPTQSPQIP	1160
QY	795	-----	794
Db	1161	NYDLNRAIPTDVSOIFPSLELVWQTATSLDLSQPSISPDLGQMALSPDPQESLSPDL	1220
QY	795	-----	794
Db	1221	QQTSLSPDLQESLSPDLGQTALSPDPQESLSPDLGQTALSPDPQESLSPDLGQTALS	1280
QY	795	-----	794
Db	1281	PDQGESLSPDLGQTSLSPDLQESLSPDLGQTALSPDPQESLSPDLGQTALSPDPQES	1340
QY	795	-----	794
Db	1341	SLSPDLGQTSLSPDLGQESLSPDLGQTALSPDPQESLSPDLGQTSLSPDLGQESLSPDL	1400
QY	795	-----	794
Db	1401	QQTALSPDLQESLSPDLGQTPLSPDLQESLSPDLQESLSPDLQESLSPDLQESLSPDL	1460
QY	795	-----	794
Db	1461	QSLPLPEFGQTFFPNADIGQMPSPDSTLNTFTFPEEPNPLVWGLSRDGDGYIEIIPRQ	1520
QY	795	-----RSFQKTKRHYFAAVER	811
Db	1521	KEESSEEDYGEFEFVAYNDPYQTLRLTDINSSRNPDIAMWLRNTGNRKRYTAAABEI	1580
QY	812	LWDYGMSSSPHVLNRRAQSGS---VPQ---PKVVFQETDGSFTQPLVYRGELNEHLGLL	865

```
Db 1581 SWDS-----KVSDDVDYVPEYVYKVKVFRKYLSDSTFTKLPDQSEYEHGIL 1631
Qy 866 GPYIRAEVEDIMVTFRNOARPSYFSLISYE-----EDORQGAERPKNFVKPNE 917
Db 1632 GPVIRAEVDDVQVFRKNSLRPSYLAHGLSYKSSRGKTYEDDSDPEWFKEDNAIQNK 1691
Qy 918 TKTYFWKHQWMAPTKDFEDCKAMAYFSDVDLEKDVHGLGTPLLVCHVTNINPAHQVQ 977
Db 1692 TTYVYHATTRSGPENPGSACRAWAYYSAVNPEDKHSLGLPLLCIKGLTLDKTNVP 1751
Qy 978 TVQEFALPFTLFDKTSVYFTEENMERNCRAPCNIQMEDPTKENYRFAHNGYMDTLP 1037
Db 1752 DMREFVLLFWFDEKSYIDKKPTRSWRRASS-----EYNSHEFAHNGMIYN-LPG 1804
Qy 1038 LVMAQDQRIWYLSMGSNENIHSIFSGHVFTRVKEEYKMAVNLNYPGVFTEMLPS 1097
Db 1805 LRMVQEWMVRLHLLNLGSRDHHVHFHGTLLNGTQHQQLGVWPLPGSFKLEMCAS 1864
Qy 1098 KAGIWRVECLIGEHLAGHAGTSLFLVYSNKQCTPLGMSGHIRDPOITASGOYQWAPKLA 1157
Db 1865 KPGWMLLTVGEIQRAGMOTPFLLVDRCKMPGLSTGLIADSIQASBFWGYWEPKLA 1924
Qy 1158 RLHYSGSINAW-----STK-BPFSWKVYDILLAPMIIHGKIQGARQKFSLYISQILMY 1211
Db 1925 RLNGSGSNAMIAEKLSTEFNPEPWIQDMQKEVLLTGICQGAHYLKPYITTEFCVAY 1984
Qy 1212 SLDGKKWQTYRGNSGTGLMVFNGVNDSSGIKHNIFNPPIIARYIRLHPHTHYSIRSLAME 1271
Db 1985 SLDRKNWIRFGKNSRNVWYFGNSDASTIKENQIDPPVARYIRISPTGSYNKPALRE 2044
Qy 1272 LMGCILNCSPLMGESKAIQASQITASSYFTNNPAT-WSPSKARLHIOGSNAWRPOVN 1330
Db 2045 LQCEVNGCSTPLMGESKIKENQITASSFKSMWNGYWEFPFLARLNAQGRVNAQAKAN 2104
Qy 1331 NPKELQVDFQTKWKTGVTGQVKSLLTSMVKEFLSSSQDGHQWTLFQNGKV--KV 1398
Db 2105 NNQWLQIDLLKIKITAITVQGCKSLSSEMYVKSYYTHYSDQGDWPKPYREKSGMWDKI 2164
Qy 1389 FQGNQDSFTPVVNSLDPPLRLYRIHPQSWHOIALRMEVLGCEAODLY 1438
Db 2165 FEGNNVYRGHKNFNPPIISFRIRIIPKTNQSIARLELFGC---DMY 2211

RESULT 6
T42764
C:Species: Mus musculus (house mouse)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jun-2000
C:Accession: T42764
R:Yang, T.L.; Cui, J.; Rehumtulla, A.; Yang, A.; Moussalli, M.; Kaufman, R.J.; Ginsburg,
Blood 91, 4593-4599, 1998
A:Title: The structure and function of murine factor V and its inactivation by protein C
A:Reference number: 222270; MUID:98282202; PMID:9616155
A:Accession: T42764
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2183 <YAN>
A:Cross-references: EMBL:U52925; NID:G3219690; PID:G3219691; PIDN:AAC99553.1
C:Function:
A:Pathway: blood coagulation
C:Superfamily: coagulation factor V; discoidin I amino-terminal homology; ferroxidase re
C:Keywords: blood coagulation; duplication; glycoprotein; phospholipid binding; plasma;
F:350-662/Domain: ferroxidase repeat homology <FOX1>
F:1541-1864/Domain: ferroxidase repeat homology <FOX2>

Query Match 30.8%; Score 2369; DB 2; Length 2183;
Best Local Similarity 25.8%; Pred. No. 1.7e-149;
Matches 577; Conservative 287; Mismatches 486; Indels 888; Gaps 31;

Qy 3 RRYLGAVELSWDYMQSDGLGELPVADARPPRPVPSFFPNTSVVYKTLFVEFTVHLFNIA 62
Db 32 RQFYAAQGLIWNHYE-----PTDPSLNSIPSPFKKIYREYEQY-FKKE 75
```

```
Qy 63 KPRPPMGLIGPTTOAEVYDVTITLKNMASHPVSLHAGVSYWKAASGAEYDDQTSORE 122
Db 76 KPRSSNSGLLPTLYAEVGVIVKHFRNKADKPLSIHPQGIKYSKFSGSAVDHTPPAE 135
Qy 123 KEDDKVPGGSHYVWVULKENGPMASDPLCLTYSLYSHVDLVKDLNSGLIGALLVCREG 182
Db 136 RKDAPVAGEEYTYEWIVSESDGPTDDPPCLTHIYYSYENLTQDFNSGLIGPLICKKG 195
Qy 183 SLAKEKTQL--HKFILLFAVFDGKSWHSETKNSLMQDRDAASARAWPKMHTVNGVNR 240
Db 196 TLTEGGTQKMDKXHVLLFAVDESKS-----RSQSPSLMYTINGFVNK 239
Qy 241 SLPLGICHRRKSVYVHVGMTTPEVHSIFLEGTFLVRNHRQASLEISPTFTTAOTLL 300
Db 240 TMPDITVCAHDHVSMLHIGMSSGPELFSIHFGQVLEQNHQKVTVTLSATSTANMTM 299
Qy 301 MDLQQLFLPCHISHQHDMGEAYVKVDSCEPEPOLRMKNNEAEADYDDDLTDSMDVVRP 360
Db 300 SPGRWIVSSLIPIKHYQAGMQAYIDIKNCPKCTR----- 333
Qy 361 DDDNSPSFQIRSVAKKHPKMTWVHIAAEEDWDYAPLVLPADDRSVKSOVLNNGPORIG 420
Db 334 ---SPK--TLTREORRYMKWEYFIAAEVINYAPVIPANMDKIYRSQHLDFNSQIG 387
Qy 421 RYKKVRFMAYTDETF--KTREAIQHSIGILGPLYGEVGTLLIIFKNQASRPYNIYPH 478
Db 388 KHYKKVYRQYBEETFTKRTDNPSIKQSGILGPVIRAQVRDLKAIKFNKASRPYSIYPH 447
Qy 479 GITDVRPLSRRLPKGV-----KHLKDPILPGEIPIKYKWTVTVEDGPTKSPRCLTR 531
Db 448 GVT-----FSPYEDGINSSTSGSHTTIRPVQPGFTTYKWNILEFDEPTENDAQCILTR 501
Qy 532 YSSFVNMERDLASGLIGPLLCYKESVDQKGNQIMSKRNVILFSYVDENRSMYLTENI 591
Db 502 PYSDVDVTRDIASGLIGLLIICKRSLDQGVORVADIEQQAFAVAFDENKSWYIEDNI 561
Qy 592 QRLFNPAGVQLEDPFQASNIMHSINGYVDSIQ--LSVCLHEVAYVYILSIGAOTPLS 650
Db 562 NKCPENDEVKRDDPKFVESNIMSTINGYVPESITLGFCDPTVQMHFCSVGHDDILT 621
Qy 651 VFGSGYFKHKWYEDTLTLPFSGGTFVMSMNPGLWILGCHNSDFNRQMTALLKVSS 710
Db 622 IHFTGHSFIYGRREDTLTLPFMRGESVTVMQNVGTWMLTTMNSPKRNLRLRFRDVK 681
Qy 711 CDKNTGDY-VEDSYE-----DISAYLLS----- 732
Db 682 CNR---DYDNEDSYEIEPPAPTSMTTRIHDSLENEFGIDNEDDDYQYLLASSLGRSF 738
Qy 733 KNAIEPR----- 740
Db 739 KNSLNPEENEFNLTALENSSEFISPTDVRVDSNSSRLSKIINNKLKDFORTLPGS 798
Qy 741 ----- 740
Db 799 GATVAGTLRLNLIGLDFNFVINSSTHRSSTSYHENDMENPQSNITWVLLPLGPKSGNR 858
Qy 741 ----- 740
Db 859 EQDKPKTIKTRGPHMKHRSWMKAPAGKTRHNSPNKNSYSGMKSEBEDIPELIPLKQKI 918
Qy 741 ----- 740
Db 919 TSKFLNRRWRVASEKGSYEIIAANGEDTDVCLKLTNSPQONQITVPRGESTSHNTTRKPS 978
Qy 741 ----- 740
Db 979 DLPTFSGVGHKSPHVRQEENSGFKQKLFTRTKKKKKKKLALHSLSPRGPDLRGN 1038
Qy 741 ----- 740
Db 1039 HSPFPDRLLNHSLLLHKSNETALSPDLNQTSPSMSTRSLPDYNOYSKNDTEQMSSLD 1098
```

```
QY 741 -----SFSQNPVVKR----- 751
Db 1099 LYQSVPAEHSPTTFAQDPDQTHSTTDPYRSPPELSQGLDYDLSDHPYDDIGLTSFF 1158
QY 752 ----- 751
Db 1159 PQSQKSPSSDDQAIPSSDLSLFTISPDLQTIIPDLQLLSPEDNQKTSPPDLQ 1218
QY 752 ----- 751
Db 1219 VPLSPDDNQKTSPPDLQGVLSLPPDDNQKTSPPDLQGVLSLDDNQKTSPPDLQVPLSPD 1278
QY 752 -----HQR-----EIRTTTISQDEB----- 767
Db 1279 DNQMITSPDLQGVPLSSDNQKTSPPDLQGVPLFPEDNQVYFLDSQVPLSSDNQKTSST 1338
QY 768 ----- 767
Db 1339 DLLTLSPDPQTVLSPDLQLPLPSDNQVTVSPDLSSLTLSPDFNEIILAPDLQVTLIS 1398
QY 768 ----- 767
Db 1399 PDLIQTNPALNHGKASSADPPQASYPDPGQASSLPFLNRLPHPLPHTHIPPSPSPPL 1458
QY 768 -----IDYDDT-----ISVMKKEDPDYDEDENQSPRSFQ 798
Db 1459 NNTSLSRKENPLVVGLSRVDGDDVEIIVPSEPERIDEDYAEDDFVYNDPYRTDTRDV 1518
QY 799 KTRH-----YFTAAVERLWDYGMSSSPHVLNRAQSGVPO-----FK 837
Db 1519 NSSRPDPTIAAWYLGRGHGKKFYIAEEITWYAEFAQSEM--DHEDTGTHTKDTYTK 1576
QY 838 KVFQFETDGSQTPLYRGLNEHLGLGPIYRAVEDNIMVTFRQASRPYSFVSSLIS 897
Db 1577 KVFYKYLSTSTPRDPRAYEHEHLGILGFVIRAEVDDVIQVRFKMLASRPYSLHAHGLS 1636
QY 898 YE-----EPQRCQAEPRKPKVFNKTKTYFWKQVHMAPTQDEDFCKAWAYSDDVDL 949
Db 1637 YKESSEKTYEDESPEMFQEDDAQVNPSSYTVWHTKSGENPSCACRANAYYSANV 1696
QY 950 EKDVHSGILGPLLVCHTNTLNPAGHQVQTVQSFALFTTFIDETKSYWYFTENMERNCRAPC 1009
Db 1697 ERDIHSGILGPLLICRKTGLHMERNLPMDBREFVLLFWFDEKKSWYKES--KGSR--- 1751
QY 1010 NQMEDPTKENVRFHAIINGYIMDTLPGLVMAQDQIRWYLLSMGSENIHHSCHVF 1069
Db 1752 --RIBSPEEKNAHKFYAINGMIYN-LPGLRMVEQEWVRLHLLNMGGSRDIHVHFGQTL 1808
QY 1070 TVRKKEBYKVALYNLYPGVFETVEMLPKSKAGIWRVECLICEHLHAGMSTFLVYSNKQCT 1129
Db 1809 LDRNRYKHQHLGVWPLLPGSFKTLEMKASKPGWLLDTEVGENQVAGMQTPFLIIDKECKM 1868
QY 1130 PLGMASGHTRDPQITASGYGOWAPKLARLHYSGSINAWSTKE-----PFS-WIKVDLLA 1183
Db 1869 PMGLSTGVLSDSQIKASELYLTWEPLRLNAGSYNWSIEKTLDFDPKEWIOVDMQK 1928
QY 1184 PMIHGIKTQBARQKSFSLYSOFITMYSIDGKQVYTGNSGTGLMVFFGNVDSGKIH 1243
Db 1929 EYVVTGIOQGAHKYLLKSCFTTFEQVAYSDDQTNWQIFRGKSGKSYMYFTGNSDGTIKE 1988
QY 1244 NTFNPPIIARYTLRHPHTHYSIRSLRMELMGCDLNSCPLMGESKAIQDQITASSYFT 1303
Db 1989 NRLDPPIVARYIRIHFTKSYNRTFLIELOGCEVNGCSTPLGLEDGRIQDKQITASSFPK 2048
QY 1304 NMFAT--WSPSKARLHLQGRSNARWQVNNPKLEQLVDFOKTMKVTVGTTQGVKSLTSMY 1362
Db 2049 SWMGDWYEPESLARLNAQGRVNAQKANNKQMLQVLDLLKIKVTAIVTQGCCKSLSEMY 2108
QY 1363 VKEFLLSSQDGHQWTLFQNGKV--KVFQGNQDSFTPVVNSLDPLLLRYLRIHPQSVW 1420
Db 2109 VKSYSIQYSDQGVAMKPYRKQKSMVDKIFEGNSNTKGMKNFFNPPIISRFTIRIIPKTN 2168
QY 1421 HQIALRMEVLGCEAQDLY 1438
```

```
Db 2169 QSIARLELPGC---DIY 2183

RESULT 7
KUUH
ferroxidase (EC 1.16.3.1) precursor [validated] - human
N:Alternate names: ceruloplasmin
N:Contains: ferroxidase long form (CP-1); ferroxidase short form (CP-2)
C:Species: Homo sapiens (man)
C:Date: 31-Aug-1980 #sequence revision 12-May-1995 #text_change 08-Dec-2000
C:Accession: A24165; A35450; A00524; I59067
R:Koshinsky, M.L.; Funk, W.D.; van Oost, B.A.; MacGillivray, R.T.A.
Proc. Natl. Acad. Sci. U.S.A. 83, 5086-5090, 1986
A:Title: Complete cDNA sequence of human ceruloplasmin.
A:Reference number: A25443; MUID:86259737; PMID:2873574
A:Accession: A25443
A:Molecule type: mRNA
A:Residues: 1-1060,1065-1069 <KOS>
A:Cross-references: GB:M13699; NID:g180255; PIDN:AAA51976.1; PID:g180256
A:Note: this is the short or CP-2 alternatively spliced form
R:Mercer, J.F.B.; Grimes, A.
FEBS Lett. 203, 185-190, 1986
A:Title: Isolation of a human ceruloplasmin cDNA clone that includes the N-terminal 1
A:Reference number: A24165; MUID:86275241; PMID:3755405
A:Accession: A24165
A:Molecule type: mRNA
A:Residues: 1-40;549-599;784-829;919-952 <MER>
R:Yang, F.; Friedrichs, W.E.; Cupples, R.L.; Bonifacio, M.J.; Sanford, J.A.; Horton,
J. Biol. Chem. 265, 10780-10785, 1990
A:Title: Human ceruloplasmin. Tissue-specific expression of transcripts produced by a
A:Reference number: A35450; MUID:90285218; PMID:2355023
A:Accession: A35450
A:Molecule type: DNA
A:Residues: 1007-1064 <YAN>
A:Cross-references: GB:J05506
A:Note: this is the long or CP-1 alternatively spliced form
R:Takahashi, N.; Ortel, T.L.; Putnam, F.W.
Proc. Natl. Acad. Sci. U.S.A. 81, 390-394, 1984
A:Title: Single-chain structure of human ceruloplasmin: the complete amino acid sequen
A:Reference number: A00524; MUID:84119493; PMID:6582496
A:Accession: A00524
A:Molecule type: protein
A:Residues: 20-1060,1065-1069 <TAK>
A:Note: 79-Gly and 449-Gly were also found
R:Yang, F.; Naylor, S.B.; Lum, J.B.; Cutshaw, S.; McCombs, J.L.; Naberhaus, K.H.; McG
Proc. Natl. Acad. Sci. U.S.A. 83, 3257-3261, 1986
A:Title: Characterization, mapping, and expression of the human ceruloplasmin gene.
A:Reference number: I59067; MUID:86205876; PMID:3486416
A:Accession: I59067
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 218-1069 <RES>
A:Cross-references: GB:M13536; NID:g180248; PIDN:AAA51975.1; PID:g180249
C:Comment: Ferroxidase is a blue, plasma alpha2-glycoprotein binding 6-7 copper ions
C:Comment: In Wilson's disease the plasma levels of ferroxidase are diminished or und
C:Comment: The three fragment chains are produced spontaneously during purification a
C:Genetics:
A:Gene: GDB:CP
A:Cross-references: GDB:119069; OMIM:117700
A:Map position: 3q23-q25
A:Introns: 1006/3; 1061/1
A:Note: the list of introns is incomplete
C:Function:
A:Description: catalyzes the oxidation of free iron(II) to iron(III) coupled with the
A:Note: iron(III), but not iron(II), is the form bound and transported by transferrin
A:Note: other possible functions are amine oxidase activity, copper transport and hom
C:Superfamily: ferroxidase; ferroxidase repeat homology
C:Keywords: acute phase; alternative splicing; copper; duplication; glycoprotein; oxi
F1-19/Domain: signal sequence #status predicted <SIG>
F20-1069/Product: ferroxidase, long form #status predicted <MATH>
F20-1060/Product: ferroxidase, short form #status experimental <MATH>
F20-499/Product: ferroxidase 67K chain #status experimental <K67>
```

**RESULT 8**

A35210  
ferroxidase [EC 1.16.3.1] precursor - rat  
N;Alternate names: ceruloplasmin  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C;Accession: A35210; A41753; A29564; S21692  
R;Fleming, R.B.; Gitlin, J.D.  
J. Biol. Chem. 265, 7701-7707, 1990  
A;Title: Primary structure of rat ceruloplasmin and analysis of tissue-specific gene expression  
A;Reference number: A35210; MUID:90237081; PMID:2332446  
A;Accession: A35210  
A;Status: not compared with conceptual translation  
A;Molecule type: mRNA  
A;Residues: 1-1059 <FLR>  
A;Cross-references: GB:J05424  
E;Fleming, R.B.; Gitlin, J.D., 1992  
J. Biol. Chem. 267, 473-486, 1992  
A;Title: Structural and functional analysis of the 5'-flanking region of the rat ceruloplasmin gene  
A;Reference number: A41753; MUID:92112697; PMID:1730611  
A;Accession: A41753  
A;Molecule type: DNA  
A;Residues: 1-48 <FL2>  
A;Note: The authors translated the codon GAA for residue 40 as Gly, GAA for residue 41 as Ala,  
R;Alfred, A.R.; Grimes, A.; Schreiber, G.; Mercer, J.F.B.  
J. Biol. Chem. 262, 2875-2878, 1987  
A;Title: Rat ceruloplasmin. Molecular cloning and gene expression in liver, choroid plexus  
A;Reference number: A29564; MUID:87137545; PMID:3818625  
A;Accession: A29564  
A;Molecule type: mRNA  
A;Residues: 'NSG', 215-216, 'Y', 218, 'FAT', 222, 'E', 224-226, 'B', 228, 'LL', 231, 'D', 233-235, 'L'  
A;Experimental source: liver  
A;Note: The authors translated the codon GCG for residue 60 as Gly and GTG for residue 61 as Val  
R;Ryan, T.P.; Grover, T.A.; Aust, S.D.  
Arch. Biochem. Biophys. 293, 1-8, 1992  
A;Title: Rat ceruloplasmin: resistance to proteolysis and kinetic comparison with human ceruloplasmin  
A;Reference number: S21692; MUID:92117681; PMID:1531003  
A;Accession: S21692  
A;Molecule type: protein  
A;Residues: 20-29, 'Q'; 502-910 <RYA>  
C;Superfamily: ferroxidase; ferroxidase repeat homology  
C;Keywords: copper; glycoprotein; oxidoreductase; plasma  
P;1-19/Domain: signal sequence #status predicted <SIG>  
P;20-1059/Product: ferroxidase #status predicted <MAT>  
P;23-356/Domain: ferroxidase repeat homology <FOI>  
P;372-712/Domain: ferroxidase repeat homology <FO2>  
P;737-1053/Domain: ferroxidase repeat homology <FO3>

**RESULT 8**

A35210  
ferroxidase [EC 1.16.3.1] precursor - rat  
N;Alternate names: ceruloplasmin  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C;Accession: A35210; A41753; A29564; S21692  
R;Fleming, R.B.; Gitlin, J.D.  
J. Biol. Chem. 265, 7701-7707, 1990  
A;Title: Primary structure of rat ceruloplasmin and analysis of tissue-specific gene expression  
A;Reference number: A35210; MUID:90237081; PMID:2332446  
A;Accession: A35210  
A;Status: not compared with conceptual translation  
A;Molecule type: mRNA  
A;Residues: 1-1059 <FLR>  
A;Cross-references: GB:J05424  
E;Fleming, R.B.; Gitlin, J.D., 1992  
J. Biol. Chem. 267, 473-486, 1992  
A;Title: Structural and functional analysis of the 5'-flanking region of the rat ceruloplasmin gene  
A;Reference number: A41753; MUID:92112697; PMID:1730611  
A;Accession: A41753  
A;Molecule type: DNA  
A;Residues: 1-48 <FL2>  
A;Note: The authors translated the codon GAA for residue 40 as Gly, GAA for residue 41 as Ala,  
R;Alfred, A.R.; Grimes, A.; Schreiber, G.; Mercer, J.F.B.  
J. Biol. Chem. 262, 2875-2878, 1987  
A;Title: Rat ceruloplasmin. Molecular cloning and gene expression in liver, choroid plexus  
A;Reference number: A29564; MUID:87137545; PMID:3818625  
A;Accession: A29564  
A;Molecule type: mRNA  
A;Residues: 'NSG', 215-216, 'Y', 218, 'FAT', 222, 'E', 224-226, 'B', 228, 'LL', 231, 'D', 233-235, 'L'  
A;Experimental source: liver  
A;Note: The authors translated the codon GCG for residue 60 as Gly and GTG for residue 61 as Val  
R;Ryan, T.P.; Grover, T.A.; Aust, S.D.  
Arch. Biochem. Biophys. 293, 1-8, 1992  
A;Title: Rat ceruloplasmin: resistance to proteolysis and kinetic comparison with human ceruloplasmin  
A;Reference number: S21692; MUID:92117681; PMID:1531003  
A;Accession: S21692  
A;Molecule type: protein  
A;Residues: 20-29, 'Q'; 502-910 <RYA>  
C;Superfamily: ferroxidase; ferroxidase repeat homology  
C;Keywords: copper; glycoprotein; oxidoreductase; plasma  
P;1-19/Domain: signal sequence #status predicted <SIG>  
P;20-1059/Product: ferroxidase #status predicted <MAT>  
P;23-356/Domain: ferroxidase repeat homology <FOI>  
P;372-712/Domain: ferroxidase repeat homology <FO2>  
P;737-1053/Domain: ferroxidase repeat homology <FO3>





E;1-21/Domain: signal sequence #status predicted <SIG>  
F;28-60/Domain: EGF homology <EG1>  
F;68-107/Domain: EGF homology <EG2>  
F;110-267/Domain: discoidin I amino-terminal homology <DN1>  
F;271-427/Domain: discoidin I amino-terminal homology <DN2>

Query Match 8.6%; Score 663; DB 2; Length 427;

Best Local Similarity 40.3%; Pred. No. 1.6e-36;

Matches 142; Conservative 61; Mismatches 123; Indels 26; Gaps 6;

QY 1105 ECLIGELHAGMSTLFLVYSNK-----CQPLGWSGHIRDFOITASGOY-- 1149

DB 78 KCLVETDQRG--DIFTYIICQCPVGSYGIEHCEGSCFKLGEGAIADSOISASSVYMG 135

QY 1150 ----GQWAPKLARLHYSGINAW--STKEPFSWKVLDLLAPMIHGIKTQARQKFSLSY 1203

DB 136 FMGLQRWGPELALRYTGIVNAWTASSYDSKFWLQVDFLRKRVSGVMTQASRAGAEY 195

QY 1204 ISQFIIMYSLDGKKWQYRGNSGTLMVFFGNVDSGKIKHINFPPIIARYIRLHPHYS 1263

DB 196 LKTEKVAISLDGRREFEIQDESQTDKEFGMGQDNNSLKINWFNFTLEAQYIRLVPSVCH 255

QY 1264 IRSLRMELMGCCLNSCMPLGMSKAISSDAQITASSYFT--NMFA-TWSPSKARLHLOG 1320

DB 256 RGCTLRFELGCEHLCSEPLGLKNTIPDSQITASSSYKTNLRAFQWYPHLGRLDNQ 315

QY 1321 RSNARFPQVNNPKWLQVDFQKTMKVGTVTQGVKSLTSMVYKEFLISSQDGHQWTLF 1380

DB 316 KINAWTAQSNASAKWLQVDTLQTKKYGIITQGRDFGHQYVASYKVAHSDGQVMTY 375

QY 1381 FQNGKVVQFQNGDSFTFVNSLDPPLTRYLRHQPQSWHQAIALRMEVLGC 1432

DB 376 BEQTSKVFQGNLNNSHKKNIFEKPFMARYVRVLPFLSWHNRITLRLLELGC 427

## RESULT 11

A36479 milk fat globule membrane protein - mouse

C;Species: Mus musculus (house mouse)

C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C;Accession: A36479

R;Stubbs, J.D.; Lekutis, C.; Singer, K.L.; Bui, A.; Yuruki, D.; Srinivasan, U.; Parry, G

Proc. Natl. Acad. Sci. U.S.A. 87, 8417-8421, 1990

A;Title: cDNA cloning of a mouse mammary epithelial cell surface protein reveals the exi

A;Reference number: A36479; MUID:91046008; PMID:2122462

A;Accession: A36479

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-463 <STU>

C;Cross-references: GB:M38337; NID:g199142; PID:AAA39534.1; PID:g199143

C;Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF homol

C;Keywords: membrane protein

F;28-60/Domain: EGF homology <EG1>

F;68-107/Domain: EGF homology <EG2>

F;147-303/Domain: discoidin I amino-terminal homology <DN1>

F;307-463/Domain: discoidin I amino-terminal homology <DN2>

Query Match 8.5%; Score 657; DB 1; Length 463;

Best Local Similarity 42.2%; Pred. No. 4.7e-36;

Matches 135; Conservative 63; Mismatches 110; Indels 12; Gaps 5;

QY 1124 SNKQTPLGWSGHIRDFOITASGOY-----GQWAPKLARLHYSGINAW--STKEPFS 1175

DB 145 ASRCSTQLGMEGGAIDASQISASYVMGFMGLQRWGPELALRYTGIVNAHASNYSK 204

QY 1176 WIKVDLLAPMIHGIKTQARQKFSLSYISQFIIMYSLDGKKWQYRGNSGTLMVFFGN 1235

DB 205 WIQVNLRLKRVSGVMTQASRAGAEYLTFTKVAISLDGRKFEFIQDESQGD-KBELGN 263

QY 1236 VDSGKIKHINFPPIIARYIRLHPHYSIRSLRMELMGCCLNSCMPLGMSKAISSDAQ 1295

DB 264 LDNNSLKNVNFNFTLEAQYIRLVPSVCHRGCTLRFELGCEHLCSEPLGLKNTIPDSQ 323

QY 1296 ITASSYFT--NMFA-TWSPSKARLHLOGRSNAMRPQVNNPKWLQVDFQKTMKVGTVTQ 1352

DB 324 MSASSSYKTNLRAFQWYPHLGRDLNQKINAWTAQSNASAKWLQVDTLQRTQVGTIIT 383

QY 1353 GVKSLTSMVYKEFLISSQDGHQWTLFFQNGKVKVQGHQSDSTFVNSLDPPLTRYL 1412

DB 384 GARDFGHIQVYVESYKVAHSDGQVMTYVBQGSKVPQGNLNNSHKKNIFEKPFMARY 443

QY 1413 RIHPQSWHQAIALRMEVLGC 1432

DB 444 RVLFPVSHNRITLRLLELGC 463

## RESULT 12

T11743

PP47 protein - pig (fragment)

C;Species: Sus scrofa domestica (domestic pig)

C;Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 08-Sep-2002

C;Accession: T11743

R;Enslin, M.; Vogel, T.; Calvete, J.J.; Thole, H.H.; Schmidtke, J.; Matsuda, T.; Toei

Biol. Reprod. 58, 1057-1064, 1998

A;Title: Molecular cloning and characterization of P47, a novel boar sperm-associated

A;Reference number: 217325; MUID:98206817; PMID:9546740

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-409 <ENS>

A;Cross-references: EMBL:Y11683; NID:g2652927; PIDN:CAA72379.1; PID:g2652928

A;Experimental source: testis

C;Function:

A;Description: may be involved in membrane remodeling and/or function as a zona pelluc

C;Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF hom

F;6-40/Domain: EGF homology <EGF>

Query Match 8.5%; Score 650; DB 2; Length 409;

Best Local Similarity 39.3%; Pred. No. 1.1e-35;

Matches 139; Conservative 62; Mismatches 125; Indels 28; Gaps 6;

QY 1105 ECLIGELHAGMSTLFLVYSNK-----CQPLGWSGHIRDFOITASGOY-- 1149

DB 58 ECEVIDDAHRG--DVFTYEICKPHGYTGHCCEIICNAPLOMETGAIDFQISASSWHLG 115

QY 1150 ----GQWAPKLARLHYSGINAW--STKEPFSWKVLDLLAPMIHGIKTQARQKFSLSY 1203

DB 116 FMGLQRWAPELARLHRAGINAWTASVNDYRNPQVNLRRMRVTGVVYQASRAGAEY 175

QY 1204 ISQFIIMYSLDGKKWQYRGNSGTLMVFFGNVDSGKIKHINFPPIIARYIRLHPHYS 1263

DB 176 MKTFKVAISTDGRKFQFIQGAEBESGDKIFMGNLNSGLKYNLFEPVLEVQVRLVPIICH 235

QY 1264 IRSLRMELMGCCLNSCMPLGMSKAISSDAQITASSYFTN--MFATWSPSKARLHLOG 1320

DB 236 RGCTLRFELGCEHLCSEPLGLKNTIPNKQITASSFTYRTWGLSAPSWFPFYARLDNQ 295

QY 1321 RSNARFPQVNNPKWLQVDFQKTMKVGTVTQGVKSLTSMVYKEFLISSQDGHQWTLF 1380

DB 296 KFNAWTAQSNASAEWLQIDLGSRQRRVTGIIITQGRDFGHQYVAAKVAISDDGVSWTEY 355

QY 1381 FQNGKVVQFQNGDSFTFVNSLDPPLTRYLRHQPQSWHQAIALRMEVLGC 1432

DB 356 RDQGALEGGKIFPGNLDNNSHKKNMFETPLTRFVRILPVAHNRITLRLLELGC 409

## RESULT 13

S65138

glycoprotein antigen MGP57/53, mammary gland - bovine (fragment)

N;Alternate names: glycoprotein component 16/major fat-globule membrane protein/MFG-E8

C;Species: Bos primigenius taurus (cattle)

C;Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 07-Aug-1998

C;Accession: S65138; G48394

R;Aoki, N.; Kishi, M.; Taniguchi, Y.; Adachi, T.; Nakamura, R.; Matsuda, T.

Biochim. Biophys. Acta 1245, 385-391, 1995

A;Title: Molecular cloning of glycoprotein antigens MGP57/53 recognized by monoclonal

A:Reference number: S65138; MUID:96125736; PMID:8541316

A:Accession: S65138

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-401 <AOK>

R:Mathier, I.H.; Banghart, L.R.; Lane, W.S.

Biochem. Mol. Biol. Int. 29, 545-554, 1993

A:Title: The major fat-globule membrane proteins, bovine components 15/16 and guinea-pig

II-like sequences.

A:Reference number: A48394; MUID:93250576; PMID:8485470

A:Accession: G48394

A>Status: preliminary

A:Molecule type: protein

A:Residues: 207-220 <MAT>

A:Experimental source: milk

A>Note: sequence extracted from NCBI backbone (NCBIP:131457)

C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF homol

C:Keywords: glycoprotein

F:1-32/Domain: EGF homology (fragment) <EG1>

F:40-79/Domain: discoidin I amino-terminal homology <DN1>

F:82-239/Domain: discoidin I amino-terminal homology <DN2>

F:243-401/Domain: discoidin I amino-terminal homology <DN2>

Query Match 8.3%; Score 635; DB 2; Length 401;

Best Local Similarity 37.9%; Pred. No. 1.1e-34;

Matches 135; Conservative 66; Mismatches 123; Indels 32; Gaps 7;

QY 1105 ECLGEHLHAGMSTLFLVYSNK-----CQTPLGNASGHIRDFQITASGQY-- 1149

DB 50 ECQVTDSDSHRG--DVFIQYICKPLGYVGHICHTCTCTSPFGMOTGAIDQSISASMHG 107

QY 1150 ----GQWAPKLARLHYSGINAWST----KEPFSWIKYDILAPMIHGIKQGRQKPS 1201

DB 108 FMGLQWAPELARLHQTGLVNAWTSQNDKNP--WIQVNLKMKWVTGVVTCGRASGA 165

QY 1202 LYISQFIIMYSLDGKKWQYRGNSTGLMVFVGNVDSGSIKHNFPPPIIARYLRHPTH 1261

DB 166 EYLKTFKVAYSTDGRQFQIQVAGRSQDKIFIGNVNSGLKINLFDTPLEQYVRLVPII 225

QY 1262 YSIRSTLRMELMGCDLNSCMPGLMESKAISSAQITASSYFTN---MFATWSPSKARLHL 1318

DB 226 CHRGTCLFELLGCELGCTEPLGKNDTIPNKQITASSYKTTWGLSAFSPFPFYARLDN 285

QY 1319 QGRSNARPOVNNPKWELQVDQKMTKVTGVTQGVKSLTSMYKEFLISSQDGHQWT 1378

DB 286 QQKFNWATQATNSASEWLIQDLGSKQKRVGTGITQGRDFGHQYVAARVAYGDDGVVTW 345

QY 1379 LFFQNG--KVKVFGQNGDSFTPVVNSLDPPLRLYLRHPSQVHQAIALRMEVLGC 1432

DB 346 EYKDPGAGESKIFPGMNDNSHKKNIFETFPQARFVRIQPVAVHNRITLRLVLLGC 401

RESULT 14

S74211

N:Alternate names: glycoprotein precursor - bovine

C:Species: Bos primigenius taurus (cattle)

C:Date: 04-Dec-1997 #sequence\_revision 12-Dec-1997 #text\_change 04-Nov-2002

C:Accession: S74211; S78114; S24181; S65138; G48394

R:Hvarregaard, J.; Andersen, M.H.; Berglund, L.; Rasmussen, J.T.; Petersen, T.E.

Eur. J. Biochem. 240, 628-636, 1996

A:Title: Characterization of glycoprotein PAS-6/7 from membranes of bovine milk fat glob

A:Reference number: S74211; MUID:97008954; PMID:8856064

A:Accession: S74211

A:Molecule type: mRNA

A:Residues: 1-427 <HVA>

A:Cross-references: EMBL:X91895; NID:g1632778; PIDN:CAA62997.1; PID:g1632779

A:Accession: S78114

A:Molecule type: protein

A:Residues: 19-85;96-110;140-165;174-216;221-232;248-277;285-293;309-337;339-420;425-427

R:Kim, D.H.; Kanno, C.; Mizokami, Y.

Biochim. Biophys. Acta 1122, 203-211, 1992

A:Title: Purification and characterization of major glycoproteins, PAS-6 and PAS-7, from

A:Reference number: S23926; MUID:92353107; PMID:1643094

A:Accession: S24181

A:Molecule type: protein

A:Residues: 383-394 <KIM>

R:Aoki, N.; Kishi, M.; Taniguchi, Y.; Adachi, T.; Nakamura, R.; Mateuda, T.

Biochim. Biophys. Acta 1245, 385-391, 1995

A:Title: Molecular cloning of glycoprotein antigens MGP57/53 recognized by monoclonal

A:Reference number: S65138; MUID:96125736; PMID:8541316

A:Accession: S65138

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 27-427 <AOK>

R:Mathier, I.H.; Banghart, L.R.; Lane, W.S.

Biochem. Mol. Biol. Int. 29, 545-554, 1993

A:Title: The major fat-globule membrane proteins, bovine components 15/16 and guinea-

II-like sequences.

A:Reference number: A48394; MUID:93250576; PMID:8485470

A:Accession: G48394

A>Status: preliminary

A:Molecule type: protein

A:Residues: 233-246 <MAT>

A:Experimental source: milk

A>Note: sequence extracted from NCBI backbone (NCBIP:131457)

C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF hom

C:Keywords: blocked amino end; disulfide bond; glycoprotein; milk

F:1-18/Domain: signal sequence #status predicted <SIG>

F:19-427/Product: PAS-6/7 protein #status experimental <MAT>

F:24-58/Domain: EGF homology <EG1>

F:66-105/Domain: discoidin I amino-terminal homology <DN1>

F:108-285/Domain: discoidin I amino-terminal homology <DN2>

F:243-427/Domain: discoidin I amino-terminal homology <DN2>

F:24-35,29-47,49-58,66-77,71-94,96-105/Disulfide bonds: #status predicted

F:27/Binding site: carbohydrate (Ser) (covalent) #status experimental

F:34/Binding site: carbohydrate (Thr) (covalent) #status experimental

F:59,227/Binding site: carbohydrate (Asn) (covalent) #status experimental

F:109-265,252-256,270-427/Disulfide bonds: #status experimental

Query Match 8.3%; Score 635; DB 2; Length 427;

Best Local Similarity 37.9%; Pred. No. 1.2e-34;

Matches 135; Conservative 66; Mismatches 123; Indels 32; Gaps 7;

QY 1105 ECLGEHLHAGMSTLFLVYSNK-----CQTPLGNASGHIRDFQITASGQY-- 1149

DB 76 ECQVTDSDSHRG--DVFIQYICKPLGYVGHICHTCTCTSPFGMOTGAIDQSISASMHG 133

QY 1150 ----GQWAPKLARLHYSGINAWST----KEPFSWIKYDILAPMIHGIKQGRQKPS 1201

DB 134 FMGLQWAPELARLHQTGLVNAWTSQNDKNP--WIQVNLKMKWVTGVVTCGRASGA 191

QY 1202 LYISQFIIMYSLDGKKWQYRGNSTGLMVFVGNVDSGSIKHNFPPPIIARYLRHPTH 1261

DB 192 EYLKTFKVAYSTDGRQFQIQVAGRSQDKIFIGNVNSGLKINLFDTPLEQYVRLVPII 251

QY 1262 YSIRSTLRMELMGCDLNSCMPGLMESKAISSAQITASSYFTN---MFATWSPSKARLHL 1318

DB 252 CHRGTCLFELLGCELGCTEPLGKNDTIPNKQITASSYKTTWGLSAFSPFPFYARLDN 311

QY 1319 QGRSNARPOVNNPKWELQVDQKMTKVTGVTQGVKSLTSMYKEFLISSQDGHQWT 1378

DB 312 QQKFNWATQATNSASEWLIQDLGSKQKRVGTGITQGRDFGHQYVAARVAYGDDGVVTW 371

QY 1379 LFFQNG--KVKVFGQNGDSFTPVVNSLDPPLRLYLRHPSQVHQAIALRMEVLGC 1432

DB 372 EYKDPGAGESKIFPGMNDNSHKKNIFETFPQARFVRIQPVAVHNRITLRLVLLGC 427

RESULT 15

A25945

coagulation factor VIII - pig (fragment)

C:Species: Sus scrofa domestica (domestic pig)

C:Date: 16-Aug-1988 #sequence\_revision 16-Aug-1988 #text\_change 21-Jul-2000

C:Accession: A25945

R:Toole, J.J.; Pittman, D.D.; Orr, E.C.; Murtha, P.; Wasley, L.C.; Kaufman, R.J.

Proc. Natl. Acad. Sci. U.S.A. 83, 5939-5942, 1986  
A;Title: A large region (approx 195 kDa) of human factor VIII is dispensable for in vitro  
A;Reference number: A25945; MUID:86287369; PMID:3016730  
A;Accession: A25945  
A;Status: nucleic acid sequence not shown; not compared with conceptual translation  
A;Molecule type: DNA  
A;Residues: 1-869 <TOO>  
C;Superfamily: coagulation factor VIII; discoidin I amino-terminal homology; ferroxidase

Query Match	7.1%	Score 543;	DB 2;	Length 869;
Best Local Similarity	75.2%	Pred. No. 5.4e-28;		
Matches 100;	Conservative 14;	Mismatches 19;	Indels 0;	Gaps 0;
QY	746	PPVLRHQRHRTTTLQSQDEEIDYDITISVEMKKEDFDIYDEDENQSPRSFQKTRHYF	805	
Db	737	PPVLRHQRDISLPTRQPEEDKMDYDDIFSTETKGEDFDIYGEDENQDPRSFQKTRHYF	796	
QY	806	IAAVERLWDYGMSSSPHVLNRAQSGSVPOFKKVFQEFDTGSGFTQPLYRGELNEHLGLL	865	
Db	797	IAAVEQLWDYGMSSSPRALNRAQNGEVPRFKKVFRRADGSFTNPSYRGELNKHGLL	856	
QY	866	GPYIRAEVEDNIM	878	
Db	857	GPYIRAEVEDNIM	869	

Search completed: December 9, 2003, 16:55:42  
Job time : 57 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 9, 2003, 16:49:51 ; Search time 18 seconds  
(without alignments)  
3756.911 Million cell updates/sec

Title: US-10-006-091-1

Perfect score: 7691

Sequence: 1 ATRRYLGAVELSWDMQSD.....VVHQIALRMEVLGCEAQDLY 1438

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	7227	94.0	2351	1 F88 HUMAN	P00451 homo sapien
2	6227	81.0	2133	1 F88_PIG	P12263 sus scrofa
3	6192	80.5	2319	1 F88_MOUSE	Q06194 mus musculus
4	2389.5	31.1	2224	1 F85 HUMAN	P12259 homo sapien
5	2372	30.8	2211	1 F85_BOVIN	Q28107 bos taurus
6	2357.5	30.7	2258	1 F85_PIG	Q96101 sus scrofa
7	1727.5	22.5	1065	1 CRU_HUMAN	P00450 homo sapien
8	1684.5	21.9	1059	1 CRU_RAT	P13635 rattus norv
9	1635.5	21.3	1062	1 CRU_MOUSE	Q61147 mus musculus
10	663	8.6	427	1 MFGM_RAT	P70490 rattus norv
11	657	8.5	463	1 MFGM_MOUSE	P21956 mus musculus
12	650	8.5	409	1 MFGM_PIG	P79385 sus scrofa
13	635	8.3	427	1 MFGM_BOVIN	Q95114 bos taurus
14	588	7.6	387	1 MFGM_HUMAN	Q08431 homo sapien
15	469.5	6.1	931	1 NRP2_HUMAN	O60462 homo sapien
16	464.5	6.0	925	1 NRP2_RAT	O35276 rattus norv
17	462.5	6.0	931	1 NRP2_MOUSE	O35375 mus musculus
18	458.5	6.0	914	1 NRP1_CHICK	P79795 gallus gall
19	451.5	5.9	922	1 NRP1_RAT	Q96799 rattus norv
20	446.5	5.8	923	1 NRP1_MOUSE	P97333 mus musculus
21	443	5.8	928	1 NRP1_XENLA	P28824 xenopus lae
22	429.5	5.6	923	1 NRP1_HUMAN	O14786 homo sapien
23	306.5	4.0	3133	1 HMC2_BOMMO	P98092 bombyx mori
24	266	3.5	764	1 CPX2_MOUSE	Q94215 mus musculus
25	261	3.4	756	1 CPX2_HUMAN	Q9436 homo sapien
26	260.5	3.4	280	1 XLR1_FUGRU	Q96785 figu rubrip
27	247	3.2	224	1 XLR1_MOUSE	Q92114 mus musculus
28	243	3.2	224	1 XLR1_HUMAN	O15537 homo sapien
29	223	2.9	722	1 CPXM_MOUSE	Q92100 mus musculus
30	217	2.8	734	1 CPXM_HUMAN	Q965m3 homo sapien
31	217	2.8	1331	1 CTA2_HUMAN	Q9unc6 homo sapien
32	211	2.7	1284	1 NFX4_DROME	Q94887 drosophila
33	208	2.7	1310	1 CTA4_MOUSE	Q99p47 mus musculus

34	206	2.7	1308	1 CTA4_HUMAN	Q9c0a0 homo sapien
35	198.5	2.6	854	1 DDR2_MOUSE	Q62371 mus musculus
36	194	2.5	1288	1 CTA3_HUMAN	Q9bz76 homo sapien
37	192	2.5	913	1 DDR1_HUMAN	Q08345 h epithelia
38	185.5	2.4	855	1 DDR2_HUMAN	Q16832 homo sapien
39	185.5	2.4	910	1 DDR1_RAT	Q63474 rattus norv
40	185.5	2.4	911	1 DDR1_MOUSE	Q93146 mus musculus
41	183	2.4	1384	1 CTA1_HUMAN	P78357 homo sapien
42	171	2.2	1381	1 CTA1_RAT	P97846 rattus norv
43	171	2.2	1385	1 CTA1_MOUSE	O54991 mus musculus
44	149.5	1.9	578	1 ASO_TORAC	Q40588 nicotiana t
45	148	1.9	622	1 YAK3_SCHPO	Q09920 schizosacch

ALIGNMENTS

RESULT 1  
F88\_HUMAN  
ID F88\_HUMAN STANDARD; PRT; 2351 AA.  
AC P00451;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Coagulation factor VIII precursor (Procoagulant component)  
DE (Antihemophilic factor) (AHF).  
GN F8 OR F8C.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP MEDLINE=6081164; PubMed=3935400;  
RA Truett M.A., Blacher R., Burke R.L., Caput D., Chu C., Dina D.,  
RA Hartog K., Kuo C.H., Masiaz F.R., Merryweather J.P., Najarian R.,  
RA Pachi C., Potter S.J., Puma J., Quiroga M., Rall L.B., Randolph A.,  
RA Urdea M.S., Valenzuela P., Dahl H.-H.M., Favalaro J., Hansen J.,  
RA Nordfang O., Ezban M.;  
RT "Characterization of the polypeptide composition of human factor  
VIII:C and the nucleotide sequence and expression of the human kidney  
cDNA.";  
RL DNA 4:333-349(1985).  
[2]  
RP MEDLINE=85061548; PubMed=6438526;  
RA Wood W.I., Capon D.J., Simonsen C.C., Eaton D.L., Gitschier J.,  
RA Keyt B., Seeburg P.H., Smith D.H., Hollingshead P., Wion K.L.,  
RA Delwart E., Tuddenham E.G.D., Vehar G.A., Lawn R.M.;  
RT "Expression of active human factor VIII from recombinant DNA clones.";  
RL Nature 312:330-337(1984).  
[3]  
RP MEDLINE=85061550; PubMed=6438528;  
RA Toole J.J., Knopf J.L., Wozney J.M., Soltzman L.A., Buecker J.L.,  
RA Pittman D.D., Kaufman R.J., Brown E., Shoemaker C., Orr E.C.,  
RA Amphlett G.W., Foster W.B., Coe M.L., Knutson G.J., Fass D.N.,  
RA Hewick R.M.;  
RT "Molecular cloning of a cDNA encoding human antihaemophilic factor.";  
RL Nature 312:342-347(1984).  
[4]  
RP MEDLINE=93265012; PubMed=1303178;  
RA Gitschier J., Wood W.I.;  
RT "Sequence of the exon-containing regions of the human factor VIII  
Hum. Mol. Genet. 1:199-200(1992).  
[5]  
RP SEQUENCE OF 2064-2070 FROM N.A.  
RA de Water N.S., Williams R., Browett P.J.;  
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
[6]  
RP SULFATION OF TYR-1699.

- RX MEDLINE=91093266; PubMed=1898735;  
 RA Leyte A., van Schijndel H.B., Niehrs C., Huttner W.B., Verbeet M.P.,  
 RA Mertens K., van Mourik J.A.;  
 RT "Sulfation of Tyr1680 of human blood coagulation factor VIII is  
 RT essential for the interaction of factor VIII with von Willebrand  
 RT factor.";  
 RL J. Biol. Chem. 266:740-746(1991).  
 RN [7]  
 RN SULFATION.  
 RP MEDLINE=92207952; PubMed=1554716;  
 RA Pittman D.D., Wang J.H., Kaufman R.J.;  
 RT "Identification and functional importance of tyrosine sulfate  
 RT residues within recombinant factor VIII.";  
 RL Biochemistry 31:3315-3325(1992).  
 RN [8]  
 RN STRUCTURE BY NMR OF 2322-2343.  
 RP MEDLINE=95200924; PubMed=7893714;  
 RA Gilbert G.E., Baleja J.D.;  
 RT "Membrane-binding peptide from the C2 domain of factor VIII forms an  
 RT amphipathic structure as determined by NMR spectroscopy.";  
 RL Biochemistry 34:3022-3031(1995).  
 RN [9]  
 RN REVIEW ON MOLECULAR BASIS OF HEMA.  
 RP MEDLINE=91221499; PubMed=1902642;  
 RA Gitschier J.;  
 RT "The molecular basis of hemophilia A.";  
 RL Ann. N.Y. Acad. Sci. 614:89-96(1991).  
 RN [10]  
 RN REVIEW ON MOLECULAR BASIS OF HEMA.  
 RP MEDLINE=89088506; PubMed=2491949;  
 RA White G.C. II, Shoemaker C.B.;  
 RT "Factor VIII gene and hemophilia A.";  
 RL Blood 73:1-12(1989).  
 RN [11]  
 RN REVIEW ON MOLECULAR BASIS OF HEMA.  
 RP MEDLINE=95245332; PubMed=7728145;  
 RA Antonarakis S.E., Kazazian H.H., Tuddenham E.G.D.;  
 RT "Molecular etiology of factor VIII deficiency in hemophilia A.";  
 RL Hum. Mutat. 5:1-22(1995).  
 RN [12]  
 RN VARIANT HEMA GLN-2326.  
 RP MEDLINE=86225434; PubMed=3012775;  
 RA Gitschier J., Wood W.I., Shuman M.A., Lawn R.M.;  
 RT "Identification of a missense mutation in the factor VIII gene of a  
 RT mild hemophiliac.";  
 RL Science 232:1415-1416(1986).  
 RN [13]  
 RN VARIANT HEMA PRO-2135.  
 RP MEDLINE=88096539; PubMed=3122181;  
 RA Levinson B., Janco R.L., Phillips J.A. III, Gitschier J.;  
 RT "A novel missense mutation in the factor VIII gene identified by  
 RT analysis of amplified hemophilia DNA sequences.";  
 RL Nucleic Acids Res. 15:9797-9805(1987).  
 RN [14]  
 RN VARIANT HEMA GLN-2228.  
 RP MEDLINE=88191889; PubMed=2833855;  
 RA Yousoufian H., Antonarakis S.E., Bell W., Griffin A.M.,  
 RA Kazazian H.H.;  
 RT "Nonsense and missense mutations in hemophilia A: estimate of the  
 RT relative mutation rate at CG dinucleotides.";  
 RL Am. J. Hum. Genet. 42:718-725(1988).  
 RN [15]  
 RN VARIANT HEMA GLY-291.  
 RP MEDLINE=88220354; PubMed=2835904;  
 RA Yousoufian H., Wong C., Aronis S., Platokoukis H., Kazazian H.H. Jr.,  
 RA Antonarakis S.E.;  
 RT "Moderately severe hemophilia A resulting from Glu-->Gly substitution  
 RT in exon 7 of the factor VIII gene.";  
 RL Am. J. Hum. Genet. 42:867-871(1988).  
 RN [16]  
 RN VARIANT HEMA CYS-1708.  
 RP MEDLINE=89274393; PubMed=2499363;  
 RA O'Brien D.P., Tuddenham E.G.;  
 RT "Purification and characterization of factor VIII 1,689-Cys: a  
 RT nonfunctional cofactor occurring in a patient with severe hemophilia  
 RT A.";  
 RL Blood 73:2117-2122(1989).  
 RN [17]  
 RN VARIANT HEMA CYS-391.  
 RP MEDLINE=90001543; PubMed=2506948;  
 RA Shima M., Ware J., Yoshioka A., Fukui H., Fulcher C.A.;  
 RT "An arginine to cysteine amino acid substitution at a critical  
 RT thrombin cleavage site in a dysfunctional factor VIII molecule.";  
 RL Blood 74:1612-1617(1989).  
 RN [18]  
 RN VARIANT HEMA LEU-189.  
 RP MEDLINE=90057680; PubMed=2510835;  
 RA Chan V., Chan T.K., Tong T.M., Todd D.;  
 RT "A novel missense mutation in exon 4 of the factor VIII:C gene  
 RT resulting in moderately severe hemophilia A.";  
 RL Blood 74:2688-2691(1989).  
 RN [19]  
 RN VARIANT HEMA LEU-2326.  
 RP MEDLINE=89197216; PubMed=2495245;  
 RA Inaba H., Fujimaki M., Kazazian H.H. Jr., Antonarakis S.E.;  
 RT "Mild hemophilia A resulting from Arg-to-Leu substitution in exon 26  
 RT of the factor VIII gene.";  
 RL Hum. Genet. 81:335-338(1989).  
 RN [20]  
 RN VARIANT HEMA HIS-391.  
 RP MEDLINE=89264602; PubMed=2498882;  
 RA Arai M., Inaba H., Higuchi M., Antonarakis S.E., Kazazian H.H. Jr.,  
 RA Fujimaki M., Hoyer L.W.;  
 RT "Direct characterization of factor VIII in plasma: detection of a  
 RT mutation altering a thrombin cleavage site  
 RT (arginine-372-->histidine).";  
 RL Proc. Natl. Acad. Sci. U.S.A. 86:4277-4281(1989).  
 RN [21]  
 RN VARIANT HEMA CYS-1708.  
 RP MEDLINE=90105723; PubMed=2104766;  
 RA Arai M., Higuchi M., Antonarakis S.E., Kazazian H.H. Jr.,  
 RA Phillips J.A. III, Janco R.L., Hoyer L.W.;  
 RT "Characterization of a thrombin cleavage site mutation (Arg 1689 to  
 RT Cys) in the factor VIII gene of two unrelated patients with  
 RT cross-reacting material-positive hemophilia A.";  
 RL Blood 75:384-389(1990).  
 RN [22]  
 RN VARIANTS HEMA GLN-2228 AND LEU-2326.  
 RP MEDLINE=90123183; PubMed=2105106;  
 RA Casula L., Murru S., Pecorara M., Ristaldi M.S., Restagno G.,  
 RA Mancuso G., Morfini M., de Biasi R., Baudo F., Carbonara A.;  
 RT "Recurrent mutations and three novel rearrangements in the factor  
 RT VIII gene of hemophilia A patients of Italian descent.";  
 RL Blood 75:662-670(1990).  
 RN [23]  
 RN VARIANT HEMA CYS-391.  
 RP MEDLINE=90329422; PubMed=1973901;  
 RA Pattinson J.K., McVey J.H., Boon M., Ajani A., Tuddenham E.G.;  
 RT "CRM+ haemophilia A due to a missense mutation (372-->Cys) at the  
 RT internal heavy chain thrombin cleavage site.";  
 RL Br. J. Haematol. 75:73-77(1990).  
 RN [24]  
 RN VARIANTS HEMA PHE-1699 AND CYS-1708.  
 RP MEDLINE=90152691; PubMed=2105906;  
 RA Higuchi M., Wong C., Kochan L., Olek K., Aronis S., Kasper C.K.,  
 RA Kazazian H.H., Antonarakis S.E.;  
 RT "Characterization of mutations in the factor VIII gene by direct  
 RT sequencing of amplified genomic DNA.";  
 RL Genomics 6:65-71(1990).  
 RN [25]  
 RN VARIANTS HEMA CYS-1728 AND ASP-1941.  
 RP MEDLINE=90169988; PubMed=2106480;  
 RA Traystman M.D., Higuchi M., Kasper C.K., Antonarakis S.E.,  
 RA Kazazian H.H.;  
 RT "Use of denaturing gradient gel electrophoresis to detect point  
 RT mutations in the factor VIII gene.";

Query Match 94.0%; Score 7227; DB 1; Length 2351;  
Best Local Similarity 61.6%; Pred. No. 0;  
Matches 1437; Conservative 0; Mismatches 1; Indels 894; Gaps 1;

Qy	1	ATRRYLGAVELSDYNDYQSDLGELPVDARPPRPVKSPFNTSVVYKTLFVBETVHLFN	60
Db	20	ATRRYLGAVELSDYNDYQSDLGELPVDARPPRPVKSPFNTSVVYKTLFVBETVHLFN	79
Qy	61	IAPRPWMLGLPTICAEVYDVTVITLKNASHPVSLHAVGVSYKASGAEYDDQTSQ	120
Db	80	IAPRPWMLGLPTICAEVYDVTVITLKNASHPVSLHAVGVSYKASGAEYDDQTSQ	139
Qy	121	REKEDDVFPFGSHTYVQVLKENGPMASDPLCTYSYLSHVDLVKDLNSGLIGALLVCR	180
Db	140	REKEDDVFPFGSHTYVQVLKENGPMASDPLCTYSYLSHVDLVKDLNSGLIGALLVCR	199
Qy	181	EGSLAKKTOTLHKFIILLFAVFDGKSWHSTKNSLMQDRDAASARAWPKMHTVNGYVNR	240
Db	200	EGSLAKKTOTLHKFIILLFAVFDGKSWHSTKNSLMQDRDAASARAWPKMHTVNGYVNR	259
Qy	241	SLPLGICHRKSVYWHVIGMGTTPVHSLFLEGTFLVRNHRQASLSIPITFLTAQTLL	300
Db	260	SLPLGICHRKSVYWHVIGMGTTPVHSLFLEGTFLVRNHRQASLSIPITFLTAQTLL	319
Qy	301	MDLQQLFLFCHISSHQHDEMEAYYKVDSCPEEPQLRMKNEEAEYDDDLTDSEMDVYRF	360
Db	320	MDLQQLFLFCHISSHQHDEMEAYYKVDSCPEEPQLRMKNEEAEYDDDLTDSEMDVYRF	379
Qy	361	DDNSPFIQIRSAVAKHPTKWTWYHIAAEBEDWDYAPLVLPADDRSKYSQVLYNNGPORIG	420
Db	380	DDNSPFIQIRSAVAKHPTKWTWYHIAAEBEDWDYAPLVLPADDRSKYSQVLYNNGPORIG	439
Qy	421	RKYKRVFMAVTDFTKTRAIQHESGTLGLLXGEVGDTLIIIFKQKASRPYNIYPHGI	480
Db	440	RKYKRVFMAVTDFTKTRAIQHESGTLGLLXGEVGDTLIIIFKQKASRPYNIYPHGI	499
Qy	481	TDVRPLYSRRLPKGVKHLKDPFILLPGEIFKFKWTVTVEDGPTKSDPRCLTRYSSFVNME	540
Db	500	TDVRPLYSRRLPKGVKHLKDPFILLPGEIFKFKWTVTVEDGPTKSDPRCLTRYSSFVNME	559
Qy	541	RDLASGLIGPLLI CYKESVDQGNQIMSDKRNVLFSVFDENRSMYLTENIORFLPNPAG	600
Db	560	RDLASGLIGPLLI CYKESVDQGNQIMSDKRNVLFSVFDENRSMYLTENIORFLPNPAG	619
Qy	601	VLEDDPFOASNIHWSINGYVFDLSQLSVCLHEVAYWYILSIGAQTDFLSVFPESGYTEKH	660
Db	620	VLEDDPFOASNIHWSINGYVFDLSQLSVCLHEVAYWYILSIGAQTDFLSVFPESGYTEKH	679
Qy	661	KWYEDTTLTLPFGSETVFMENPGLMWILGCHNSDFRNRGWTALLKVSSCDKNTGDYIE	720
Db	680	KWYEDTTLTLPFGSETVFMENPGLMWILGCHNSDFRNRGWTALLKVSSCDKNTGDYIE	739
Qy	721	DSYEDISAYLSKKNALBPRSF-----	742
Db	740	DSYEDISAYLSKKNALBPRSF-----	799
Qy	743	-----	742
Db	800	KIONVSSDLLMLLRQSPTPHGLSLDLOEAKYETSDDPSPGALDSNNSISEMTHFRPQ	859
Qy	743	-----	742
Db	860	LHSGDMVFTPESGQLRLNEKLGTTAATELKKLDFKVSSTSNNLISTIPSDNLAAGTDN	919
Qy	743	-----	742
Db	920	TSSLGPSPMPHYDQLDITLFGKSSPLTESGGPLSLSEENNDKSLLESGLMNSQESSW	979
Qy	743	-----	742
Db	980	GKNVSSBGRFLKGRAHGAPALLTKONALFKVISILLKTNKTSNNSATNRKTRIDGPSL	1039

Qy	743	-----	742
Db	1040	LIENSPSVQNIILESTEFKVKVTPLIHDRMLMDKNATALRLNEMSNKTTSSKNMEMVQOK	1099
Qy	743	-----	742
Db	1100	KEGPIPPDAQNPMSFPKMLFLPESARWQRTGKNSLNSGGQSPKQLVSLGPEKSVEG	1159
Qy	743	-----	742
Db	1160	QNFLSEKNVVGCGEFTKDVGLKEMVFPSSRNFLTNLNDLHNHTNQEKKIQEBIEK	1219
Qy	743	-----	742
Db	1220	KETLIQENVVLPOIHTVGTGTFNFMKNLFLSTRONVEGSDYAYAPVLODRSLNDSTNR	1279
Qy	743	-----	742
Db	1280	TKKHTAHFSKGBEENLEGLCNQTKQIVEKVACTTRISFNTSQNFVTOQRSKRALQKFL	1339
Qy	743	-----	742
Db	1340	PLEBTELEKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKEKGAIQTQSLDCLTRSHSI	1399
Qy	743	-----	742
Db	1400	POANRSPPIAKVSSPSPSIRPIYLRVLFDQNSSHLPAASRYRKDQSGVESHFLOGAKK	1459
Qy	743	-----	742
Db	1460	NNLSAILLTLEMTGDQREVSGSLGTSATNSVYKKVENTVLPKPDLPTSGKVLLPKVHI	1519
Qy	743	-----	742
Db	1520	YQKDLFPTETSNPGHLDLVEGSLLOGTBGAIKWNEANRPGVPLRVATESSAKTPSK	1579
Qy	743	-----	742
Db	1580	LLDPLAWNHYGTQIPKEEMKSOBSEKTAFAKKKDTILSINACESNHAIAINEGONKP	1639
Qy	743	-----SONPVLKXHQREITRTTLOSDQBEIDYDDTISVEMKKEDFDIY	786
Db	1640	EIEVTWAKQGRTERLCSQNPPLKXHQREITRTTLOSDQBEIDYDDTISVEMKKEDFDIY	1699
Qy	787	DEDENQSPRFQKKTTRHYFAAVERLWDYGMSSSPHYLRNRAQSGSVPOFKKVVVFQEFDT	846
Db	1700	DEDENQSPRFQKKTTRHYFAAVERLWDYGMSSSPHYLRNRAQSGSVPOFKKVVVFQEFDT	1759
Qy	847	GSFTQPLYRGELNEHLGLGPIYIRAEVEDNIMVTFRNOASRPYSFYSSLSIYEBDQROGA	906
Db	1760	GSFTQPLYRGELNEHLGLGPIYIRAEVEDNIMVTFRNOASRPYSFYSSLSIYEBDQROGA	1819
Qy	907	EPRKNFYKPNETKTYFWKVQHHMPTKDEPDCKAWAYPSVDLEKDVHSGLIGPLLVCHT	966
Db	1820	EPRKNFYKPNETKTYFWKVQHHMPTKDEPDCKAWAYPSVDLEKDVHSGLIGPLLVCHT	1879
Qy	967	NTLNPAHGRQVTVQEFALFFTIFDETCKSWYFTENNERNCRAPCNIQMEDPTFKENYRPHA	1026
Db	1880	NTLNPAHGRQVTVQEFALFFTIFDETCKSWYFTENNERNCRAPCNIQMEDPTFKENYRPHA	1939
Qy	1027	INGYIMDTLPGLVMAQDQRIRWYLLSMGNSNENIHSIHFSGHVFTVRKKKEEYKMALYNLYP	1086
Db	1940	INGYIMDTLPGLVMAQDQRIRWYLLSMGNSNENIHSIHFSGHVFTVRKKKEEYKMALYNLYP	1999
Qy	1087	GVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKQOTPLGMAHSHIROFQITAS	1146
Db	2000	GVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKQOTPLGMAHSHIROFQITAS	2059
Qy	1147	GOYQWAPKLARLHYSINAWSTKBPFSWIKVDLLAPMIITHGKTQAROKFSLSYISQ	1206
Db	2060	GOYQWAPKLARLHYSINAWSTKBPFSWIKVDLLAPMIITHGKTQAROKFSLSYISQ	2119
Qy	1207	FIIMYSLDGKKQWYRGNSTGTLMVFFGNVDSSGKHNIFNPPIIARYIRLHPHTYSIRS	1266



Db 2120 FIIMYSLDKKQWYRGNGSTGLMVFFGVNDUSSGKKNIFPPPIARVIRLPHYSIRS 2179  
Qy 1267 TIRMELMGCDLNSCMPGMSKAISDAQITASSYFTNMFTWSPSKARLHLQGRSNAWR 1326  
Db 2180 TIRMELMGCDLNSCMPGMSKAISDAQITASSYFTNMFTWSPSKARLHLQGRSNAWR 2239  
Qy 1327 PQVNNPKWLOVDQKTKMKYTGVTTOGVKSLTSMYKVEFLISSODGHQWTLFPQNGKV 1386  
Db 2240 PQVNNPKWLOVDQKTKMKYTGVTTOGVKSLTSMYKVEFLISSODGHQWTLFPQNGKV 2299  
Qy 1387 KVFQGNQDSFPPVNSLDPPLLTRYLRIHPQSWHVOIALRMEVLCGEAQDLY 1438  
Db 2300 KVFQGNQDSFPPVNSLDPPLLTRYLRIHPQSWHVOIALRMEVLCGEAQDLY 2351

RESULT 2  
FAS\_PIG STANDARD; PRT; 2133 AA.  
AC P12263; Q95243;  
DI 01-OCT-1989 (Rel. 12, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Coagulation factor VIII precursor (Procoagulant component).  
GN F8 OR CF8.  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
OX NCBI\_TaxID=9823;  
RN [1]\_\_  
RP SEQUENCE FROM N.A.  
RA Healey J.F., Lubin I.M., Lollar P.;  
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 705-1573 FROM N.A.  
RX MEDLINE=86287369; PubMed=3016730;  
RA Toole J.J., Pittman D.D., Orr E.C., Murtha P., Wasley L.C.,  
RA Kaufman R.J.;  
RT "A large region (approximately equal to 95 kDa) of human factor VIII  
is dispensable for in vitro procoagulant activity.";  
RL Proc. Natl. Acad. Sci. U.S.A. 83:5939-5942(1986).  
RN [3]  
RP SEQUENCE OF 392-759 FROM N.A.  
RX MEDLINE=94179260; PubMed=7510693;  
RA Lubin I.M., Healey J.F., Scandella D., Runge M.S., Lollar P.;  
RT "Elimination of a major inhibitor epitope in factor VIII.";  
RL J. Biol. Chem. 269:8639-8641(1994).  
CC -!- FUNCTION: FACTOR VIII, ALONG WITH CALCIUM AND PHOSPHOLIPID, ACTS  
AS A COPACITOR FOR FACTOR IXA WHEN IT CONVERTS FACTOR X TO THE  
ACTIVATED FORM, FACTOR XA.  
CC -!- SUBCELLULAR LOCATION: Extracellular.  
CC -!- SIMILARITY: Contains 3 F5/8 type A domains.  
CC -!- SIMILARITY: Contains 2 F5/8 type C domains.  
CC -!- SIMILARITY: STRONG, TO COAGULATION FACTOR V.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U49517; AAB06705.1; -.  
DR PIR; A25945; A25945.  
DR PIR; T42763; T42763.  
DR HSSP; P00451; 1CFG  
DR InterPro; IPR001117; Cu-oxidase.  
DR InterPro; IPR000421; FAS8\_C.  
DR Pfam; PF00394; Cu-oxidase; 3.  
DR Pfam; PF00754; F5\_F8 type C; 2.  
DR SMART; SM00231; FAS8C\_2.  
DR PROSITE; PS01285; FAS8C\_1; 2.

DR PROSITE; PS01286; FAS8C\_2; 2.  
DR PROSITE; PS00022; FAS8C\_3; 2.  
DR PROSITE; PS00079; MULTICOPPER OXIDASE1; 3.  
KW Blood coagulation; Repeat; Plasma; Acute phase; Calcium;  
KW Signal; Glycoprotein; Sulfation.  
FT SIGNAL 1 19 POTENTIAL.  
FT CHAIN 20 2133 COAGULATION FACTOR VIII.  
FT DOMAIN 20 357 F5/8 TYPE A 1.  
FT DOMAIN 20 199 PLASTOCYANIN-LIKE 1.  
FT DOMAIN 207 357 PLASTOCYANIN-LIKE 2.  
FT DOMAIN 399 730 F5/8 TYPE A 2.  
FT DOMAIN 399 730 PLASTOCYANIN-LIKE 3.  
FT DOMAIN 583 730 PLASTOCYANIN-LIKE 4.  
FT DOMAIN 760 1599 B.  
FT DOMAIN 1495 1822 F5/8 TYPE A 3.  
FT DOMAIN 1495 1659 PLASTOCYANIN-LIKE 5.  
FT DOMAIN 1669 1822 PLASTOCYANIN-LIKE 6.  
FT DOMAIN 1822 1970 F5/8 TYPE C 1.  
FT DOMAIN 1975 2127 F5/8 TYPE C 2.  
FT SITE 391 392 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).  
FT SITE 759 760 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).  
FT SITE 1449 1450 CLEAVAGE (ACTIVATION) (BY SIMILARITY).  
FT SITE 737 737 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).  
FT MOD\_RES 738 738 SULFATION (BY SIMILARITY).  
FT MOD\_RES 742 742 SULFATION (BY SIMILARITY).  
FT DISULFID 547 573 PROBABLE.  
FT DISULFID 1633 1659 PROBABLE.  
FT DISULFID 1822 1970 BY SIMILARITY.  
FT DISULFID 1975 2127 BY SIMILARITY.  
FT CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 259 259 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 601 601 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 929 929 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 985 985 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1025 1025 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1111 1111 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1181 1181 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1208 1208 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1245 1245 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1265 1265 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1335 1335 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1408 1408 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1611 1611 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1919 1919 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CONFLICT 713 713 N -> M (IN REF. 2).  
FT CONFLICT 734 734 I -> T (IN REF. 2).  
FT CONFLICT 792 792 G -> Q (IN REF. 2).  
FT CONFLICT 1133 1133 E -> F (IN REF. 2).  
FT CONFLICT 1191 1191 I -> L (IN REF. 2).  
FT CONFLICT 1209 1209 R -> F (IN REF. 2).  
FT CONFLICT 1437 1437 C -> G (IN REF. 2).  
FT CONFLICT 1456 1456 F -> R (IN REF. 2).  
FT CONFLICT 1539 1539 F -> R (IN REF. 2).  
FT CONFLICT 1546 1546 Q -> N (IN REF. 2).  
SQ SEQUENCE 2133 AA; 239304 MW; 152BBAB997F570DA CRC64;  
Query Match 81.0%; Score 6227; DB 1; Length 2133;  
Best Local Similarity 57.3%; Pred. No. 0;  
Matches 1212; Conservative 106; Mismatches 119; Indels 678; Gaps 3;  
Qy 1 ATRRYVLGAVELSDWYMOQSD-LGELPVDARPPRPVKFPFPFTSVYVKKTLPEVFTVHLF 59  
Db 20 ATRRYVLGAVELSDWYMOQSD-LGELPVDARPPRPVKFPFPFTSVYVKKTLPEVFTVHLF 79  
Qy 60 NIAPRPPPMWGLLGTPTQAEVYDVTITLKNASHPVSILHAGVSVYWKASGAEYDDQTS 119  
Db 80 SVARPPPMWGLLGTPTQAEVYDVTITLKNASHPVSILHAGVSVYWKASGAEYDDQTS 139  
Qy 120 QREKEDDKVFPQGSHTYVWVLKENGPMASDPLCTYTYSVLSHVDLVKDLNSGLIGALLVC 179  
Db 140 QREKEDDKVLPKGSQTYVWVLKENGPMASDPLCTYTYSVLSHVDLVKDLNSGLIGALLVC 199



OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6 X CBA; TISSUE=Liver;  
 RX MEDLINE=93300511; PubMed=8314577;  
 RA Elder B., Lakich D., Gitschier J.;  
 RT "Sequence of the murine factor VIII cDNA";  
 RL Genomics 16:374-379(1993).  
 CC -!- FUNCTION: FACTOR VIII, ALONG WITH CALCIUM AND PHOSPHOLIPID, ACTS  
 CC AS A COFACTOR FOR FACTOR IXA WHEN IT CONVERTS FACTOR X TO THE  
 CC ACTIVATED FORM, FACTOR XA.  
 CC -!- SUBCELLULAR LOCATION: Extracellular.  
 CC -!- TISSUE SPECIFICITY: FOUND IN MOST TISSUES.  
 CC -!- SIMILARITY: Contains 3 F5/8 type A domains.  
 CC -!- SIMILARITY: Contains 2 F5/8 type C domains.  
 CC -!- SIMILARITY: STRONG, TO COAGULATION FACTOR V.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L05573; AAA37385.1; -;  
 DR F1R; A47004; A47004.  
 DR HSP; P00451; 1CFG.  
 DR MGD; MGI:88383; F8.  
 DR InterPro; IPR001117; Cu-oxidase.  
 DR InterPro; IPR000421; FA58\_C.  
 DR Pfam; PF00394; Cu-oxidase; 3.  
 DR Pfam; PF00754; F5\_F8\_type\_C; 2.  
 DR SMART; SM00231; FA58C; 2.  
 DR PROSITE; PS01285; FA58C\_1; 2.  
 DR PROSITE; PS01286; FA58C\_2; 2.  
 DR PROSITE; PS00222; FA58C\_3; 2.  
 DR PROSITE; PS00079; MULTICOPPER\_OXIDASE1; 3.  
 DR Blood coagulation; Repeat; Plasma; Acute phase; Calcium;  
 KW Signal; Glycoprotein; Sulfation.  
 FT SIGNAL 1 19  
 FT CHAIN 20 2319  
 FT DOMAIN 20 349  
 FT DOMAIN 20 199  
 FT DOMAIN 207 349  
 FT DOMAIN 399 730  
 FT DOMAIN 399 573  
 FT DOMAIN 583 730  
 FT DOMAIN 760 1640  
 FT DOMAIN 1683 2008  
 FT DOMAIN 1683 1845  
 FT DOMAIN 1855 2008  
 FT DOMAIN 2008 2156  
 FT DOMAIN 2161 2313  
 FT SITE 391 392  
 FT SITE 759 760  
 FT SITE 1678 1679  
 FT SITE 1324 1325  
 FT SITE 1640 1641  
 FT MOD\_RES 367 367  
 FT MOD\_RES 737 737  
 FT MOD\_RES 738 738  
 FT MOD\_RES 742 742  
 FT MOD\_RES 1669 1669  
 FT MOD\_RES 1687 1687  
 FT DISULFID 173 199  
 FT DISULFID 547 573  
 FT DISULFID 1819 1845  
 FT DISULFID 2008 2156  
 FT DISULFID 2161 2313  
 FT CARBOHYD 61 61  
 FT CARBOHYD 233 233

FT CARBOHYD 259 259  
 FT CARBOHYD 423 423  
 FT CARBOHYD 601 601  
 FT CARBOHYD 880 880  
 FT CARBOHYD 958 958  
 FT CARBOHYD 1015 1015  
 FT CARBOHYD 1022 1022  
 FT CARBOHYD 1026 1026  
 FT CARBOHYD 1044 1044  
 FT CARBOHYD 1076 1076  
 FT CARBOHYD 1087 1087  
 FT CARBOHYD 1136 1136  
 FT CARBOHYD 1161 1161  
 FT CARBOHYD 1192 1192  
 FT CARBOHYD 1255 1255  
 FT CARBOHYD 1268 1268  
 FT CARBOHYD 1273 1273  
 FT CARBOHYD 1274 1274  
 FT CARBOHYD 1302 1302  
 FT CARBOHYD 1316 1316  
 FT CARBOHYD 1340 1340  
 FT CARBOHYD 1378 1378  
 FT CARBOHYD 1797 1797  
 FT CARBOHYD 2105 2105  
 FT CARBOHYD 2319 2319  
 SQ SEQUENCE 2319 AA; 266148 MW; FD054DE051DB2A01 CRC64;  
 Query Match 80.5%; Score 6192; DB 1; Length 2319;  
 Best Local Similarity 53.4%; Pred. No. 0;  
 Matches 1231; Conservative 97; Mismatches 103; Indels 876; Gaps 9;  
 QY 1 ATRRYVLGAVELSDWYMQSD-LGELPVDAPPPRPVPSPPFNTSVVYKTLFVFTVHLF 59  
 DB 20 ATRRYVLGAVELSDWYMQSD-LGELPVDAPPPRPVPSPPFNTSVVYKTLFVFTVHLF 79  
 QY 60 NIAKPRPPWGLLGPTIOAEVYDVTITLKNMASHPVSLHAGVSVWKASEGAYDDQTS 119  
 DB 80 NIAKPRPPWGLLGPTIOAEVYDVTITLKNMASHPVSLHAGVSVWKASEGAYDDQTS 139  
 QY 120 QREKEDKYPGSGSHYVWOLKENGPMASDPLCTYSVLSHVDVLDKLSGLIGALLVC 179  
 DB 140 QREKEDKYPGSGSHYVWOLKENGPMASDPLCTYSVLSHVDVLDKLSGLIGALLVC 199  
 QY 180 REGSLAKEKTQTLHKFILLFAVDFECKSWHSETKNSLMQDRDAASARAWPKMTVNGYN 239  
 DB 200 REGSLAKEKTQTLHKFILLFAVDFECKSWHSETKNSLMQDRDAASARAWPKMTVNGYN 259  
 QY 240 RSLPGLIGCHRSVYVHWI GMGTPPEVHSIFLEGTFLVRNHRQASLEISPTFLTAQTL 299  
 DB 260 RSLPGLIGCHRSVYVHWI GMGTPPEVHSIFLEGTFLVRNHRQASLEISPTFLTAQTL 319  
 QY 300 LMDLGQFLAFCHTSSHHQDGMAYVVDSCPEPQLRMK-NNEEAEDYDDDLTDSMDVV 358  
 DB 320 LMDLGQFLAFCHTSSHHQDGMAYVVDSCPEPQLRMK-NNEEAEDYDDDLTDSMDVV 378  
 QY 359 RFDDNSPSFIQIRSVAKKHPTWVHYIAAEEDWDYAPLVAPDERSYKSQYLNNGPOR 418  
 DB 379 TLDYDSSP-FIQIRSVAKKHPTWVHYIAAEEDWDYAPLVAPDERSYKSQYLNNGPOR 437  
 QY 419 IGRYKVKRPMATDFTKTRAIQHESGLIGPLLYGEVGDITLLIFKQASRPYNIYPH 478  
 DB 438 IGRYKVKRPMATDFTKTRAIQHESGLIGPLLYGEVGDITLLIFKQASRPYNIYPH 497  
 QY 479 GITDVPYLRRLPKGVKHLKOPPLIPGRTFKYKWTVTVEDGPTKSDPRCLTYSSFFVN 538  
 DB 498 GITDVPYLRRLPKGVKHLKOPPLIPGRTFKYKWTVTVEDGPTKSDPRCLTYSSFFVN 557  
 QY 539 MERDLASGLIGPLLYCYKESVDQRGNQIMSDKENVILFSVFDENRSMYLTENTQRELPNP 598  
 DB 558 MERDLASGLIGPLLYCYKESVDQRGNQIMSDKENVILFSVFDENRSMYLTENTQRELPNP 617  
 QY 599 AGVQLEDPEFQASINMHSINGVYVDFSLSQLSVCLHEVAYVWILSIGAQDTFLSFFSGYTF 658  
 DB 618 AKTQPDGPGFQASINMHSINGVYVDFSLSQLSVCLHEVAYVWILSIGAQDTFLSFFSGYTF 677



RT of internal repeats.";  
 RL Biochemistry 26:6508-6514 (1987).  
 RN [4]  
 RP SEQUENCE OF 1188-1215 AND 1315-2224 FROM N.A.  
 RX MEDLINE=86313665; PubMed=3092220;  
 RA Kane W.H., Davie E.W.;  
 RT "Cloning of a cDNA coding for human factor V, a blood coagulation  
 factor homologous to factor VIII and ceruloplasmin.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 83:6800-6804 (1986).  
 RN [5]  
 RP PARTIAL SEQUENCE FROM N.A.  
 RX TISSUE=Fibroblast;  
 RC MEDLINE=93203619; PubMed=8454869;  
 RA Shen N.L.L., Fan S.-T., Pyati J., Graff R., Lapolla R.J.,  
 RT Edgington T.S.;  
 RL "The serine protease cofactor factor V is synthesized by  
 lymphocytes.";  
 RT J. Immunol. 150:2992-3001 (1993).  
 RN [6]  
 RP SULFATION.  
 RX MEDLINE=94264012; PubMed=8204629;  
 RA Pittman D.D., Tomkinson K.N., Michnick D., Selighsohn U.,  
 RT Kaufman R.J.;  
 RL "Posttranslational sulfation of factor V is required for efficient  
 thrombin cleavage and activation and for full procoagulant activity.";  
 RL Biochemistry 33:6952-6959 (1994).  
 RN [7]  
 RP SULFATION.  
 RX MEDLINE=90366699; PubMed=2168225;  
 RA Hortin G.L.;  
 RT "Sulfation of tyrosine residues in coagulation factor V.";  
 RL Blood 76:946-952 (1990).  
 RN [8]  
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 2065-2224.  
 RX MEDLINE=20052169; PubMed=10586886;  
 RA Macedo-Ribeiro S., Bode W., Huber R., Quinn-Allen M.A., Kim S.W.,  
 RT Ortel T.L., Bourenkov G.P., Bartunik H.D., Stubbs M.T., Kane W.H.,  
 RA Fuentes-Prior P.;  
 RL "Crystal structures of the membrane-binding C2 domain of human  
 coagulation factor V.";  
 RL Nature 402:434-439 (1999).  
 RN [9]  
 RP VARIANT MET-1764.  
 RX MEDLINE=95179146; PubMed=7874144;  
 RA Bayston T.A., Ireland H., Olds R.J., Thein S.L., Lane D.A.;  
 RT "A polymorphism in the human coagulation factor V gene.";  
 RL Hum. Mol. Genet. 3:2085-2085 (1994).  
 RN [10]  
 RP VARIANT APCR GLN-534.  
 RX MEDLINE=94217810; PubMed=8164741;  
 RA Bertina R.M., Koelmen B.P.C., Koster T., Rosendaal F.R.,  
 RA Dirven R.J., de Ronde H., van der Velden P.A., Reitsma P.H.;  
 RT "Mutation in blood coagulation factor V associated with resistance to  
 activated protein C.";  
 RL Nature 369:64-67 (1994).  
 RN [11]  
 RP VARIANTS ILE-1285 AND ARG-1327.  
 RX MEDLINE=96351768; PubMed=8713778;  
 RA Lunghi B., Iacoviello L., Gemmati D., Dilasio M.G., Castoldi E.,  
 RA Pinotti M., Castaman G., Redaelli R., Mariani G., Marchetti G.,  
 RA Bernardi F.;  
 RT "Detection of new polymorphic markers in the factor V gene:  
 association with factor V levels in plasma.";  
 RL Thromb. Haemost. 75:45-48 (1996).  
 RN [12]  
 RP VARIANT APCR GLY-334, AND VARIANT LYS-513.  
 RX MEDLINE=98122763; PubMed=9454741;  
 RA Chan W.P., Lee C.K., Kwong Y.L., Lam C.K., Liang R.;  
 RT "A novel mutation of Arg306 of factor V gene in Hong Kong Chinese.";  
 RL Blood 91:1135-1139 (1998).  
 RN [13]  
 RP VARIANT APCR THR-334.  
 RX MEDLINE=98122764; PubMed=9454742;

RA Williamson D., Brown K., Luddington R., Baglin C., Baglin T.;  
 RT "Factor V Cambridge: a new mutation (Arg306-to-Thr) associated with  
 resistance to activated protein C.";  
 RL Blood 91:1140-1144 (1998).  
 RN [14]  
 RP VARIANTS HIS-107; THR-413; LYS-513; SER-809; THR-817; ARG-858;  
 RP ARG-865; GLU-925; GLN-1146; ALA-1530; SER-1685; VAL-1749; MET-1764;  
 RP ILE-1820 AND GLY-2222, AND VARIANT APCR GLN-534.  
 RX MEDLINE=99318093; PubMed=10391209;  
 RA Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N.,  
 RA Shaw N., Lane C.R., Lim E.P., Kalyanaraman N., Nemes J., Ziaugra L.,  
 RA Friedland L., Rolfe A., Warrington J., Lipshutz R., Daley G.Q.,  
 RA Lander E.S.;  
 RT "Characterization of single-nucleotide polymorphisms in coding regions  
 of human genes.";  
 RL Nat. Genet. 22:231-238 (1999).  
 RN [15]  
 RP ERRATUM.  
 RA Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N.,  
 RA Shaw N., Lane C.R., Lim E.P., Kalyanaraman N., Nemes J., Ziaugra L.,  
 RA Friedland L., Rolfe A., Warrington J., Lipshutz R., Daley G.Q.,  
 RA Lander E.S.;  
 RL Nat. Genet. 23:373-373 (1999).  
 CC -!- FUNCTION: COAGULATION FACTOR V IS A COFACTOR THAT PARTICIPATES  
 CC WITH FACTOR XA TO ACTIVATE PROTHROMBIN TO THROMBIN.  
 CC -!- SUBUNIT: Factor Va is composed of a heavy chain and a light  
 CC chain, noncovalently bound. The interaction between the two chains  
 CC is calcium-dependent.  
 CC -!- DOMAIN: DOMAIN B CONTAINS 35 X 9 AA TANDEM REPEATS, AND 2 X 17 AA  
 CC REPEATS.  
 CC -!- PTM: Thrombin activates factor V proteolytically to the active  
 CC cofactor, factor Va (formation of a heavy chain at the N-  
 CC terminus and a light chain at the C-terminus).  
 CC -!- PTM: SULFATION IS REQUIRED FOR EFFICIENT THROMBIN CLEAVAGE AND  
 CC ACTIVATION AND FOR FULL PROCOAGULANT ACTIVITY.  
 CC -!- DISEASE: Defects in F5 are the cause of Owren parahemophilia  
 CC [MTM:227400], an hemorrhagic diathesis.  
 CC -!- DISEASE: Defects in F5 are the cause of resistance to activated  
 CC protein C (APCR) [MIM:188055], a form of thrombophilia. The APCR  
 CC mutation is found in about 5% of the population which suggest that  
 CC a slight thrombotic tendency may confer some advantage in fetal  
 CC implantation.  
 CC -!- SIMILARITY: Contains 3 F5/8 type A domains.  
 CC -!- SIMILARITY: Contains 2 F5/8 type C domains.  
 CC -!- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L32779; AAB59401.1; JOINED.  
 DR EMBL; L32755; AAB59401.1; JOINED.  
 DR EMBL; L32756; AAB59401.1; JOINED.  
 DR EMBL; L32757; AAB59401.1; JOINED.  
 DR EMBL; L32758; AAB59401.1; JOINED.  
 DR EMBL; L32759; AAB59401.1; JOINED.  
 DR EMBL; L32760; AAB59401.1; JOINED.  
 DR EMBL; L32761; AAB59401.1; JOINED.  
 DR EMBL; L32762; AAB59401.1; JOINED.  
 DR EMBL; L32763; AAB59401.1; JOINED.  
 DR EMBL; L32764; AAB59401.1; JOINED.  
 DR EMBL; L32765; AAB59401.1; JOINED.  
 DR EMBL; L32766; AAB59401.1; JOINED.  
 DR EMBL; L32767; AAB59401.1; JOINED.  
 DR EMBL; L32768; AAB59401.1; JOINED.  
 DR EMBL; L32769; AAB59401.1; JOINED.  
 DR EMBL; L32770; AAB59401.1; JOINED.  
 DR EMBL; L32771; AAB59401.1; JOINED.  
 DR EMBL; L32772; AAB59401.1; JOINED.







FT	DOMAIN	1748	1890	PLASTOCYANIN-LIKE 6.	
FT	DOMAIN	1894	2048	F5/8 TYPE C 1.	
FT	DOMAIN	2053	2208	F5/8 TYPE C 2.	
FT	SITE	741	742	CLEAVAGE (BY THROMBIN)	(BY SIMILARITY)
FT	SITE	1034	1035	CLEAVAGE (BY THROMBIN)	(BY SIMILARITY)
FT	SITE	1564	1565	CLEAVAGE (BY THROMBIN)	(BY SIMILARITY)
FT	DISULFID	167	193	PROBABLE.	
FT	DISULFID	499	525	PROBABLE.	
FT	DISULFID	1712	1738	PROBABLE.	
FT	DISULFID	1894	2048	BY SIMILARITY.	
FT	DISULFID	2053	2208	BY SIMILARITY.	
FT	MOD RES	697	697	SULFATION (POTENTIAL).	
FT	MOD RES	701	701	SULFATION (POTENTIAL).	
FT	MOD RES	730	730	SULFATION (POTENTIAL).	
FT	MOD RES	1513	1513	SULFATION (POTENTIAL).	
FT	MOD RES	1529	1529	SULFATION (POTENTIAL).	
FT	MOD RES	1537	1537	SULFATION (POTENTIAL).	
FT	MOD RES	1541	1541	SULFATION (POTENTIAL).	
FT	CARBOHYD	225	225	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	239	239	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	297	297	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	382	382	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	460	460	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	553	553	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	587	587	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	745	745	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	756	756	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	774	774	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	780	780	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	902	902	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	952	952	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	964	964	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1044	1044	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1053	1053	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1062	1062	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1071	1071	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1078	1078	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1094	1094	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1451	1451	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1490	1490	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1550	1550	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1690	1690	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1839	1839	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1997	1997	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	2196	2196	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	VARIANT	587	592	NFTLPA -> T (IN VARIANT 2).	
SQ	SEQUENCE	2211	AA; 248981	MM; CBBF90B738667C45	CRC64;

Query Match  
Best Local Similarity 30.8%; Score 2372; DB 1; Length 2211;  
Matches 594; Conservative 265; Mismatches 487; Indels 924; Gaps 35;

Qy	3	RRYLGAVELSWDYNQSDGLBGLVDARPPRPVPSKFPNTSVYVKKTLFVETVHLFNIA	62
Db	32	RQFYAAQSIKRWNYR-----PESTHL-----SSKPPETS--PKIVIREYRAY-FQKE	76
Qy	63	KRPPEWGLLPTTOAEVYDVTWTLKMMASHPVSLHAGVSYWKASEGAEYDDOTSORE	122
Db	77	KPQERTSGLLPTLYAEVGDIMKVFKNKAHPUSIHAQGIKYSKFSGASYSHTLPME	136
Qy	123	KEDDKVPFGGSHYTVVQVLKENGPMASDPLCLITYSLYSHVDLVKDLNSGLIGALLVCREG	182
Db	137	KWDDAVAPQBYTYEWIISSEHSGPHDDPPCLTHIYYSVNLVEDFNSGLIGPLICKG	196
Qy	183	SLAKEKTTL--HKFILLFVDFDEKGSWHSEKNSLMODRDAASARAWPKMHTVNGYNR	240
Db	197	TLTSDGTOKMFEKQHVLMFAVEDSKSNQTS-----LMTVNGYVNG	240
Qy	241	SLPLGLGHRKSVVWHVIGMTYEVHSIFLEGHTFLVNRHQASLESPITFTLTAQLL	300
Db	241	TMPDITVCAHDHISWHLGSSGPELFSIFHFNQGVLEQNHHKISAITLVSATSTTANMTV	300
Qy	301	MDLQGLFLFCHISSHQHDMGMEAYVKVDSCEPPEQLRMKNNEEAEDYDDDLTSDMDVVR	360

Db	301	SPEGRWTIASLIPRHFOAGQVAYIDHKNCAKTRNPKK-----LTRDQ-----	343
Qy	361	DDNSPSFQIRSVAKKPKTKWVHYIAAEEDMDYAPLVLPADDDRSYKSOVLNNGPORIG	420
Db	344	-----RRHIKRWEYFIAAEVINDYAFIIPANMDKKTRSLHLDNFSNRIG	388
Qy	421	RKYKVRPMAYTDTFKTR--EAIQHESGILGPLLYGEVGDTLIIIFNQAISREYNYPH	478
Db	389	KHYKVVYQYQDDSFTRKLEDDSSGDLGILPIIRAQVDRDTLAKIVFNKMASRSYSIVPH	448
Qy	479	GIT-----DVRPLYSRRLPKGVKHLKDFPILPGEIPKYKWTVTVEGPTKSDRCL	529
Db	449	GVTFSPYDNEVNSSTSGSNTMIRAV-----PGETTYKWNILSEDEPTENDACQL	500
Qy	530	TRYSSFWNMRDLASGLIGLICYKESVDQGNQIMSKRNVLFSVPDENRSWYLTE	589
Db	501	TRPYSNVDITRDLASGLIGLICKRSIDRRGIQRAADIEQQAQVAVFVDEKNSWTIED	560
Qy	590	NIQRFLPNPAGVQLEDPEFOASNMH-----SINGYVFDLSQ--LSVCLHEVAYWYILSIG	643
Db	561	NIYKFCENPEKVRKDDPKFYESNIMSNFTLPAINGYVPESIPILGFCFDDTVQHFCSVG	620
Qy	644	AQTDPLSVFSGYTFKHKMYVEDTLTLFPFSGETVFMENPGLWILGCTHNSDPRNGMT	703
Db	621	TONDILTIHFTGHSFYIKRHEDTLTLFPMQGESVTVMNVGTWMLTWNNSPRSKKL	680
Qy	704	ALLKVVSCDKNTGD-----YYEDS-----YEDISAYLL-----	731
Db	681	LRFDACKIRNDDDDSYEIIYFSGSTAMTKKHDSIEDENDADSDYQDELALILGL	740
Qy	732	-----SKNNAISPRFSQ-----	744
Db	741	RSPFNSSINQKDELNTALALEKDSFIPPSANRSLDSNSSSRSHVSRUJAKNFASLX	800
Qy	745	-----NPPVLK-----	750
Db	801	TLHLLEAPAGSPLEHAGLKGKNGALNPPMAEHSSPYSEDREDHPLSDVTGVSLLPFGT	860
Qy	751	-----RHOR-----ETRTTLQSDQ-----EEI-----	768
Db	861	FKNRKAHQHOFQVGRGQAQKHFSQTRFPAHKTRTLSDQNSSSRMGPMEDIPSDLL	920
Qy	769	-----	768
Db	921	LQKQDPYKILNGEWHLYSEKSYEIIODANENKTVNKLPSNPQSDSRTWGENIFPKNSHG	980
Qy	769	-----	768
Db	981	KQSGHPTFLVTRRKPLQDRDRNRSLKEGLPLIRTRKKKEKPAYHVPLSPRSFHLR	1040
Qy	769	-----DYDD-----	772
Db	1041	GEVNASFSDRRHNSLLHLLHASNETSIDLNQTFPSMNLASALPDHDDQSPNDTTSQT	1100
Qy	773	-----TISVENKEDFDIYDED-----ENOSP-----	794
Db	1101	SSPDLTYTVPSEHYQIFPIQSDPHTHTTAPSNRSPDPHTHTTAPSNRSPDPHTTAP	1160
Qy	795	-----	794
Db	1161	NYDLNRALPTDVSOIIPFSLVLEWQATSLDLSQPSISPDLGQWALSPPDQESLSPDL	1220
Qy	795	-----	794
Db	1221	QOTSLSPLSQESLSPDLGQTALSPPDPSQESLSPDLGQTALSPPDPSQESLSPDLGQTALS	1280
Qy	795	-----	794
Db	1281	PDPQESLSPDLGQTALSPPDPSQESLSPDLGQTALSPPDPSQESLSPDLGQTALSPPDPSQ	1340
Qy	795	-----	794

Db 1341 SLSPDLGQTSPLDGLQBSLSPLDGLQALSPDPQBSLSPLDGLQTSPLDGLQBSLSPLD 1400  
 QY 795 ----- 794  
 Db 1401 GQALSPDLQBSLSPLDGLQTSPLDGLQBSLSPLDGLQALSPDPQBSLSPLDGLQTSPLDGLQBSLSPLD 1460  
 QY 795 ----- 794  
 Db 1461 QSLPLPFGQTFNADIGQMPSPDPSTLNTWFIPEFNPVLVVGSLRDDGDIYELIPRQ 1520  
 QY 795 -----RSQKTRHYFAAVER 811  
 Db 1521 KEESSEEDYGEFEVAYNDPYQTLRTDINSRNPDAIAWYLRGNTGRKYYIAASEI 1580  
 QY 812 LWDYGMSSPHVLRNRAQSGS---VPO---FKKVVQBFQTSFTQPLRYRGELNHLGL 865  
 Db 1581 SWDYS-----RFVQSDDDYVPEFTVYKVVFRKYLDTFTKLPDQGEYEHGL 1631  
 QY 866 GPVIRAEVDNIMVTFRNOASRPYSFYSLISYE-----EDQRQGAEPKRNFKVNE 917  
 Db 1632 GPVIRAEVDNIMVTFRNOASRPYSFYSLISYE-----EDQRQGAEPKRNFKVNE 1691  
 QY 918 TKTYFMKVOHMAPTKDEFCWAYFSDVDLEKDVHSLGLGLPLVCHTNTLNPAGRQV 977  
 Db 1692 TVTYVWHATTRSGPENPGSACWAYSAVNPKEIDHSLGLGLPLVCHTNTLNPAGRQV 1751  
 QY 978 TVQEFALFTTIDETKSNVTENMRNCRAPCNQIOMEDTFFKYNRFAHNGYIMDTLPG 1037  
 Db 1752 DMREFVLLFMVDEKSNWYDKKTRSWRRASS-----EVKNSHEPFAHNGYIMYIN-LPG 1804  
 QY 1038 LVMAQDQIRWLLSGNSNENIHSFSGHVFTVRKKEKYNALNLYPGVETVEMLPS 1097  
 Db 1805 LEMVQEWVRLHLLNGLGSRDLHVVFHFGQTLLENGTQHQHGLVWPLLPSPKTLKMAS 1864  
 QY 1098 KAGIWRVECLIGEHLHAGMSTLFLVYNNKCPPLGWSAGHIRDFQITASGQYQWAPKLA 1157  
 Db 1865 KPGWLLDTEVGEIQRAGQTFPLIVDRECKPMGLSTGLIADSIQASEFNGYWEFKLA 1924  
 QY 1158 RLHYSGSINAW-----STK-EFSPWIKVDLLAPMIHGTQAGKQKFSLSYISQIFIMY 1211  
 Db 1925 RLNNGSGYNAMTAELSTEFNPEPIQVDMQKVELLGTQAGKHYLKPYYTTFECVAY 1984  
 QY 1212 SLDGKWKQYRNGSTGLTVFPGNDVSGIKENINFPPIIARYIKLHPHYISRTLRME 1271  
 Db 1985 SLDRKNWRLFKGNSTNVMVFGNSDASTIKENQIDPPVARYIISPTGSKYKALRL 2044  
 QY 1272 LMGCDLNCSMPLGMESKAISDAQITASSYFTNMFAT-WSPSKARHLHQGRNAPQVN 1330  
 Db 2045 LQCEVNGCGSTFLGMESGKIENKQITASSFKKSWGNYWEPFLARLNAQGRVNAQAKAN 2104  
 QY 1331 NPKWLOVDFOKTMKVGTGTQGVKSLLTSMYKVELISSQDGHQWTLFFQNGKV--KV 1388  
 Db 2105 NNQWLOVDLKLKIKITAIVTQCKSLSEMYYKSYTHYSDQGTDMKPYREKSNVVDKI 2164  
 QY 1389 FQGNODSPVNVSLDPLLTLYRLTHPQSWHQAIRMEVLGCEAQDLY 1438  
 Db 2165 FEGNNVRGHVKNFNPPIISIRIIPKTNQSIALLRLELPGC----DMY 2211

## RESULT 6

FA5\_PIG STANDARD; PRT; 2258 AA.  
 AC O9GLE1,  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Coagulation factor V precursor (Activated protein C cofactor).  
 GN F5.  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OX NCBI\_TaxID=9823;  
 RN [1]

RP SEQUENCE FROM N.A., AND 3D-STRUCTURE MODELING OF F5/8 TYPE A AND C  
 RP DOMAINS.  
 RC TISSUE=Liver;  
 RX MEDLINE=21121490; PubMed=11229814;  
 RA Grimm D.R., Colter M.B., Braunschweig M., Alexander L.J., Neame P.J.,  
 RA Kim H.K.W.;  
 RT "Porcine factor V: cDNA cloning, gene mapping, three-dimensional  
 RT protein modeling of membrane binding sites and comparative anatomy of  
 RT domains." Life Sci. 58:148-159(2001).  
 RL CC -!- FUNCTION: Coagulation factor V is a cofactor that participates  
 CC with factor Xa to activate prothrombin to thrombin.  
 CC CC -!- SUBUNIT: Factor Va is composed of a heavy chain and a light  
 CC chain, noncovalently bound. The interaction between the two chains  
 CC is calcium-dependent.  
 CC CC -!- DOMAIN: Domain B contains 41 X 9 AA tandem repeats. Domains C1  
 CC and C2 may be involved in membrane binding.  
 CC CC -!- PTM: Thrombin activates factor V proteolytically to the active  
 CC cofactor, factor Va (formation of a heavy chain at the N-  
 CC terminus and a light chain at the C-terminus).  
 CC CC -!- SIMILARITY: Contains 3 F5/8 type A domains.  
 CC CC -!- SIMILARITY: Contains 2 F5/8 type C domains.  
 CC CC -!- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; AF191308; AAG28381.1; --  
 DR HSP5; P12259; 1CZT.  
 DR InterPro; IPR001117; Cu-oxidase.  
 DR InterPro; IPR000421; FA58\_C.  
 DR Pfam; PF00394; Cu-oxidase; 3.  
 DR Pfam; PF00754; F5\_F8\_type\_C; 2.  
 DR SMART; SM00231; FA58C\_2; 2.  
 DR PROSITE; PS01285; FA58C\_1; 2.  
 DR PROSITE; PS01286; FA58C\_2; 2.  
 DR PROSITE; PS00022; FA58C\_3; 2.  
 DR PROSITE; PS00079; MULTICOPPER\_OXIDASE1; 2.  
 DR Blood coagulation; Glycoprotein; Sulfation; Calcium; Signal; Zymogen;  
 KW Repeat.  
 KW SIGNAL.  
 FT CHAIN 1 22  
 FT CHAIN 23 2258  
 FT CHAIN 23 737  
 FT PEPTIDE 738 1611  
 FT -----  
 FT CHAIN 1612 2258  
 FT DOMAIN 30 329  
 FT DOMAIN 30 193  
 FT DOMAIN 203 329  
 FT DOMAIN 348 683  
 FT DOMAIN 348 525  
 FT DOMAIN 535 683  
 FT DOMAIN 691 1611  
 FT DOMAIN 1168 1539  
 FT REPEAT 1168 1176  
 FT REPEAT 1177 1185  
 FT REPEAT 1186 1194  
 FT REPEAT 1195 1203  
 FT REPEAT 1204 1212  
 FT REPEAT 1213 1221  
 FT REPEAT 1222 1230  
 FT REPEAT 1231 1239  
 FT REPEAT 1240 1248  
 FT REPEAT 1249 1257  
 FT REPEAT 1258 1266  
 FT REPEAT 1267 1275  
 FT REPEAT 1276 1284

41 X 9 AA APPROXIMATE TANDEM REPEATS OF  
 T-L-S-P-D-L-[GS]-[HQ]-T.





RT "Single-chain structure of human ceruloplasmin: the complete amino-  
acid sequence of the whole molecule.";  
RL Proc. Natl. Acad. Sci. U.S.A. 81:390-394(1984).  
RN [6]  
RP SEQUENCE OF 158-333; 518-724 AND 858-1065.  
RX MEDLINE=83117800; PubMed=6571985;  
RA Takahashi N., Bauman R.A., Orlow T.L., Dwulet F.E., Wang C.-C.,  
RN Putnam F.W.;  
RT "Internal triplication in the structure of human ceruloplasmin.";  
RL Proc. Natl. Acad. Sci. U.S.A. 80:115-119(1983).  
RN [7]  
RP SEQUENCE OF 501-905.  
RX MEDLINE=81199407; PubMed=6940148;  
RA Dwulet F.E., Putnam F.W.;  
RN "Complete amino acid sequence of a 50,000-dalton fragment of human  
ceruloplasmin.";  
RL Proc. Natl. Acad. Sci. U.S.A. 78:790-794(1981).  
RN [8]  
RP SEQUENCE OF 907-1065.  
RX MEDLINE=80137543; PubMed=6987229;  
RA Kingston I.B., Kingston B.L., Putnam F.W.;  
RN "Primary structure of a histidine-rich proteolytic fragment of human  
ceruloplasmin. I. Amino acid sequence of the cyanogen bromide  
peptides.";  
RL J. Biol. Chem. 255:2878-2885(1980).  
RN [9]  
RP SEQUENCE OF 907-1065.  
RX MEDLINE=80137544; PubMed=6987230;  
RA Kingston I.B., Kingston B.L., Putnam F.W.;  
RN "Primary structure of a histidine-rich proteolytic fragment of human  
ceruloplasmin. II. Amino acid sequence of the tryptic peptides.";  
RL J. Biol. Chem. 255:2886-2896(1980).  
RN [10]  
RP SEQUENCE OF 1007-1061 FROM N.A.  
RX MEDLINE=90285218; PubMed=2355023;  
RA Yang F.M., Friedrichs W.E., Cupples R.L., Banifacio M.J.,  
RN Sanford J.A., Horton W.A., Bowman B.H.;  
RT "Human ceruloplasmin. Tissue-specific expression of transcripts  
produced by alternative splicing.";  
RL J. Biol. Chem. 265:10780-10785(1990).  
RN [11]  
RP REVIEW.  
RX MEDLINE=22049919; PubMed=12055353;  
RA Hellman N.E., Gitlin J.D.;  
RN "Ceruloplasmin metabolism and function.";  
RT Annu. Rev. Nutr. 22:439-458(2002).  
RN [12]  
RP X-RAY CRYSTALLOGRAPHY (3.1 ANGSTROMS).  
RA Zaitseva I., Zaitsev V., Card G., Moshkov K., Bax B., Ralph A.,  
RN Lindley P.;  
RT "The X-ray structure of human serum ceruloplasmin at 3.1 A: nature of  
the copper centres.";  
RL J. Biol. Inorg. Chem. 1:15-23(1996).  
CC -1- FUNCTION: CERULOPLASMIN IS A BLUE, COPPER-BINDING (6-7 ATOMS PER  
MOLECULE) GLYCOPROTEIN FOUND IN PLASMA. FOUR POSSIBLE FUNCTIONS  
ARE FERROXIDASE ACTIVITY, AMINE OXIDASE ACTIVITY, COPPER TRANSPORT  
AND HOMEOSTASIS, AND SUPEROXIDE DISMUTASE ACTIVITY.  
CC -1- CATALYTIC ACTIVITY: 4 Fe(2+) + 4 H(+) + O(2) = 4 Fe(3+) + 2 H(2)O.  
CC -1- COFACTOR: BINDS 6 CU-IONS PER MOLECULE. THIS PROTEIN BELONGS TO  
THE MULTICOPPER OXIDASES WHICH CONTAIN THREE DISTINCT CU CENTERS  
KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE 3 OR COUPLED  
BINUCLEAR.  
CC -1- TISSUE SPECIFICITY: SYNTHESIZED IN LIVER AND SECRETED INTO THE  
PLASMA.  
CC -1- DISEASE: Defects in CP are the cause of aceruloplasminemia  
[MIM:604290], an autosomal recessive disorder of iron metabolism.  
CC It is characterized by iron accumulation in the brain as well as  
visceral organs. Clinical features consist of the triad of retinal  
degeneration, diabetes mellitus and neurological disturbances.  
CC -1- DISEASE: Ceruloplasmin levels are decreased in Wilson's disease,  
in which copper cannot be incorporated into ceruloplasmin in  
liver because of defects in the copper-transporting ATPase 2.  
CC -1- SIMILARITY: Contains 3 F5/8 type A domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M13699; AAA51976.1; --  
DR EMBL; D45045; BAA08085.1; --  
DR EMBL; D45044; BAA08084.1; --  
DR EMBL; D45028; BAA08084.1; JOINED.  
DR EMBL; D45029; BAA08084.1; JOINED.  
DR EMBL; D45030; BAA08084.1; JOINED.  
DR EMBL; D45031; BAA08084.1; JOINED.  
DR EMBL; D45032; BAA08084.1; JOINED.  
DR EMBL; D45033; BAA08084.1; JOINED.  
DR EMBL; D45034; BAA08084.1; JOINED.  
DR EMBL; D45035; BAA08084.1; JOINED.  
DR EMBL; D45036; BAA08084.1; JOINED.  
DR EMBL; D45037; BAA08084.1; JOINED.  
DR EMBL; D45038; BAA08084.1; JOINED.  
DR EMBL; D45039; BAA08084.1; JOINED.  
DR EMBL; D45040; BAA08084.1; JOINED.  
DR EMBL; D45041; BAA08084.1; JOINED.  
DR EMBL; D45042; BAA08084.1; JOINED.  
DR EMBL; D45043; BAA08084.1; JOINED.  
DR EMBL; D00025; BAA00019.1; --  
DR EMBL; X04135; CAA27752.1; --  
DR EMBL; X04136; CAA27753.1; --  
DR EMBL; X04137; CAA27754.1; --  
DR EMBL; X04138; CAA27755.1; --  
DR EMBL; M13536; AAA51975.1; --  
DR EMBL; J05506; -- NOT\_ANNOTATED\_CDS.  
DR PIR; A25443; KUH0.  
DR PDB; 1KCW; 12-FEB-97.  
DR GlycoSuiteDB; P00450; --  
DR SWISS-2DPAGE; P00450; HUMAN.  
DR Slena-2DPAGE; P00450; --  
DR Genew; HGNC:2295; Cp.  
DR MLM; 117700; --  
DR MIM; 604290; --  
DR GO; GO:0005615; C:extracellular space; TAS.  
DR GO; GO:0006879; P:iron ion homeostasis; TAS.  
DR InterPro; IPR001117; Cu-oxidase.  
DR InterPro; IPR002355; MultiCu-oxidase2.  
DR Pfam; PF00394; Cu-oxidase; 3.  
DR PROSITE; PS00079; MULTICOPPER OXIDASE1; 3.  
DR PROSITE; PS00080; MULTICOPPER OXIDASE2; 1.  
DR Oxidoreductase; Copper; Metal-binding; Glycoprotein; Plasma; Repeat;  
KW Signal; Polymorphism; 3D-structure.  
FT SIGNAL 1 19  
FT CHAIN 20 1065 CERULOPLASMIN.  
FT F5/8 TYPE A 1.  
FT DOMAIN 20 357 PLASTOCYANIN-LIKE 1.  
FT DOMAIN 20 200 PLASTOCYANIN-LIKE 2.  
FT DOMAIN 209 357 PLASTOCYANIN-LIKE 2.  
FT DOMAIN 370 718 F5/8 TYPE A 2.  
FT DOMAIN 370 560 PLASTOCYANIN-LIKE 3.  
FT DOMAIN 570 718 PLASTOCYANIN-LIKE 4.  
FT DOMAIN 730 1061 F5/8 TYPE A 3.  
FT DOMAIN 730 900 PLASTOCYANIN-LIKE 5.  
FT DOMAIN 908 1061 PLASTOCYANIN-LIKE 6.  
FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .).  
FT CARBOHYD 358 358 N-LINKED (GLCNAC. . .).  
FT CARBOHYD 397 397 N-LINKED (GLCNAC. . .).  
FT CARBOHYD 762 762 N-LINKED (GLCNAC. . .).  
FT DISULFID 174 200 PROBABLE.  
FT DISULFID 276 357 PROBABLE.  
FT DISULFID 534 560 PROBABLE.  
FT DISULFID 637 718 PROBABLE.  
FT DISULFID 874 900 PROBABLE.  
FT METAL 120 120 COPPER (TYPE 2) (BY SIMILARITY).







AC Q61147;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Ceruloplasmin precursor (EC 1.16.3.1) (Ferroxidase).  
 GN Cp.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Klomp L.W.J., Farhangrazi Z.S., Choi D.W., Gitlin J.D.;  
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP TISSUE SPECIFICITY.  
 RP MEDLINE=96294736; PubMed=8690795;  
 RA Klomp L.W.J., Farhangrazi Z.S., Dugan L.L., Gitlin J.D.;  
 RT "Ceruloplasmin gene expression in the murine central nervous system.";  
 RL J. Clin. Invest. 98:207-215 (1996).  
 CC -!- FUNCTION: CERULOPLASMIN IS A BLUE, COPPER-BINDING (6-7 ATOMS PER  
 CC MOLECULE) GLYCOPROTEIN FOUND IN PLASMA. FOUR POSSIBLE FUNCTIONS  
 CC ARE FERROXIDASE ACTIVITY, AMINE OXIDASE ACTIVITY, COPPER TRANSPORT  
 CC AND HOMEOSTASIS, AND SUPEROXIDE DISMUTASE ACTIVITY.  
 CC -!- CATALYTIC ACTIVITY: 4 Fe(2+) + 4 H(+) + O(2) = 4 Fe(3+) + 2 H(2)O.  
 CC -!- COFACTOR: BINDS 6 CU-IONS PER MOLECULE. THIS PROTEIN BELONGS TO  
 CC THE MULTICOPPER OXIDASES WHICH CONTAIN THREE DISTINCT CU CENTERS  
 CC KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE 3 OR COUPLED  
 CC BINUCLEAR.  
 CC -!- TISSUE SPECIFICITY: MANY TISSUES, INCLUDING LIVER, EYE AND BRAIN.  
 CC -!- SIMILARITY: Contains 3 F5/8 type A domains.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 DR HMBL; U49430; AAB07996.1; -;  
 DR HSSP; P00450; 1KCW.  
 DR MGD; MGI:88476; Cp.  
 DR InterPro; IPR001117; Cu-oxidase.  
 DR InterPro; IPR002355; MultiCu\_oxidase2.  
 DR Pfam; PF00394; Cu-oxidase; 3.  
 DR PROSITE; PS00079; MULTICOPPER\_OXIDASE1; 3.  
 DR PROSITE; PS00080; MULTICOPPER\_OXIDASE2; 1.  
 KW Oxidoreductase; Copper; Metal-binding; Glycoprotein; Plasma; Repeat;  
 KW Signal.  
 FT SIGNAL. 1 19  
 FT CHAIN 20 1062  
 FT DOMAIN 20 356  
 FT DOMAIN 20 199  
 FT DOMAIN 208 356  
 FT DOMAIN 269 713  
 FT DOMAIN 369 555  
 FT DOMAIN 565 713  
 FT DOMAIN 725 1057  
 FT DOMAIN 725 896  
 FT DOMAIN 904 1057  
 FT DISULFID 173 199  
 FT DISULFID 275 356  
 FT DISULFID 529 555  
 FT DISULFID 632 713  
 FT DISULFID 870 896  
 FT METAL 120 120  
 FT METAL 122 122  
 FT METAL 179 179  
 FT METAL 181 181  
 FT METAL 990 990  
 FT METAL 993 993  
 FT METAL 995 995  
 FT METAL 995 995

FT METAL	1035	1035	COPPER (TYPE 3) (BY SIMILARITY).
FT METAL	1036	1036	COPPER (TYPE 1) (BY SIMILARITY).
FT METAL	1037	1037	COPPER (TYPE 3) (BY SIMILARITY).
FT METAL	1041	1041	COPPER (TYPE 1) (BY SIMILARITY).
FT METAL	1046	1046	COPPER (TYPE 1) (BY SIMILARITY).
FT CARBOHYD	138	138	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	226	226	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	583	583	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	625	625	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	757	757	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	922	922	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE	1062 AA;	121159 MW;	F3F52ED09A238F16 CRC64;

Query Match 21.3%; Score 1635.5; DB 1; Length 1062;  
 Best Local Similarity 32.4%; Pred. No. 4.8e-99;  
 Matches 376; Conservative 195; Mismatches 438; Indels 153; Gaps 24;

QY 3 RRYTLGAVELSDYMQ--SDLGELPVDARFPVRVPKSPFNSTVVYKTLFVFTVHLFN 60	
DB 22 KHYFIGITEAVNDVAGSTEEKLISVDETEQSNFYLONGPDRIGRYKALFEYTDGTF 81	
QY 61 IAKRPPMGLLPTQAEVYDVTITLKNMASHPVSLHVGVSYWKASGEAYDDQTS 120	
DB 82 KTIIDKPAWGLFGLPVKAEVEDKVVVHLKSLASRIYTFHAGVTVTKEYEGAVPDNTD 141	
QY 121 REKEDDKVPGSGSHYVWQVLEKNGPMASDPLCLTYSLVSHVDLVKDLNSLIGALLVCR 180	
DB 142 FORADDDKLPQOQYVYVHLHA-NEPSPGEGSDNCVTRIYHSHVDPAKPTASGLIGLIICK 200	
QY 181 EGSLLAKEKQTQL-HKFIILLFAVDFBGKSWHSETKSLM-----QDRDAASARAWPKMHT 233	
DB 201 KGSLYKEKEKIDDEFVLMFVSVDENLSWYLEDNIKIFCSEPEKVDKDNEDFQBSNRMY 260	
QY 234 VNGVYNSRLPGLICHRKSVYVHVGMTTPEHSIFLEGHTFLVRHROASLESPITP 293	
DB 261 INGYTFGSLPGLSCAADRVKWLFGMGNEVDVSHAFPHGQALTSRNYQTDIINLFPATL 320	
QY 294 LTAQTLLMDLQOFLFCHISSHQDGMAYVYKVDSCPEEPQIRMKNEAEADYDDDLTDS 353	
DB 321 IDAYVAQNPQVWMLSCQNLNHLKAGLQAFQVQDC-----NKPSKD----- 363	
QY 354 EMDVVRFDNDSPSFIQIRSVAKKPKTWHVIAAEEDWDYAP--LVLA-PDDR-----S 406	
DB 364 -----NIRGKHVRH-----YIIAAEVIWNYAFSGIDIFTEEKLTAGSG 402	
QY 407 YKSOVLNNGPQIRGKYKVRFMAYTDETF---KTREAIQHESGILGPLYGEVDTLII 463	
DB 403 DSGVFFEQCATRIGGSYKMAVREYTDGTFNRKBERGPDEHGLGILGSPVIAEVDGTTKV 462	
QY 464 IFKQASRPYNIYPHGIT---DVRPLYSRRLPKGVKHLKDFP-ILPGEIFKYKWTVTVE 518	
DB 463 TFRNGQHLSIQPMGVSFTAEENGTYG---PPGASSQQAASHVAPKKTFTYETVTPKE 519	
QY 519 DGPTKSDPRCLTRYSSFPNNMERDLASGLIGPLITCYKESVDQGNQIMSKRNVILPSV 578	
DB 520 MGPTVADPVCLSKMYSAVDPTKDIFTGLIGPMKICKGSLADGRQKVDVDFYLPFTV 579	
QY 579 FDENRSWYLTNIORFLPNPAGVOLEDPPEFOASNMHSINGYVDSLQ--SVCLHEVAYW 637	
DB 580 FDENESILLDDNIRMFTAPQVDKEDDFQESNKNHSMNGFMVYGNQSWPHMCLGESIVW 639	
QY 638 YLSIGAGTDFLSVFFSGYTFKHKVYEDTLTLPPPSGETVFMENMNPGLMILCHNSDF 697	
DB 640 YLFSAGNEADVHGIFYSGNTYLCGBERDTANLFPKSLTLLMNPDTKGTDFVECLTDDH 699	
QY 698 RNRGWTALLKVSSCDKNTGDYEDSYEDISAYLLSKNNAIEPRSFQSNPPVLKRQREIT 757	
DB 700 YTGWKQKYTVNQCR-----QPEDFTYILGER----- 727	
QY 758 RTLLQSDQEEIDYDDTISVEMKKEDFDIYDDEBNQSPRSFQKTRHYPIAAVERLWDYGM 817	
DB 728 --TYVDAVEWMDYSPSRAWEKELHL--QEQVNSVFLDK--EFFIGS----- 772	





DR HSP; P00740; 1EDM.  
 DR InterPro; IPR006209; EGF like.  
 DR InterPro; IPR000421; FA58\_C.  
 DR InterPro; IPR006210; IEGF.  
 DR Pfam; PF00008; EGF; 2.  
 DR Pfam; PF00754; F5\_F8\_type\_C; 2.  
 DR SMART; SM00181; EGF\_2.  
 DR SMART; SM00231; FA58C; 2.  
 DR PROSITE; PS00022; EGF\_1; 2.  
 DR PROSITE; PS01186; EGF\_2; 2.  
 DR PROSITE; PS01285; FA58C\_1; 2.  
 DR PROSITE; PS01286; FA58C\_2; 2.  
 DR PROSITE; PS00022; FA58C\_3; 2.  
 KW Glycoprotein; Repeat; EGF-like domain.  
 FT DOMAIN 2 41  
 FT DOMAIN 44 88  
 FT DOMAIN 91 247  
 FT DOMAIN 252 409  
 FT SITE 67 69  
 FT SITE CELL ATTACHMENT SITE (POTENTIAL).  
 FT DISULFID 6 17  
 FT DISULFID 11 29  
 FT DISULFID 31 40  
 FT DISULFID 91 247  
 FT DISULFID 234 238  
 FT DISULFID 252 409  
 FT CARBOHYD 41 41  
 FT CARBOHYD 372 372  
 SQ SEQUENCE 409 AA; 45725 MW; BOC07AF80029927A CRC64;  
 Query Match 8.5%; Score 650; DB 1; Length 409;  
 Best Local Similarity 39.3%; Pred. No. 3.7e-35;  
 Matches 139; Conservative 62; Mismatches 125; Indels 28; Gaps 6;  
 QY 1105 ECLIGELHAGMSTFLVYSNK-----COTPLGMSGHIRDFQITASGOY-- 1149  
 Db 58 ECEVIDAHRG--DVFTEYCKPHGVGTGHCIIICNAPMETGATDFQISSAHLG 115  
 QY 1150 ----GOWAPKLARLHYSGSINAW--STKEPFSWIKVDLLAPMIITHGKTQCAROKFSSLY 1203  
 Db 116 FMGLQRAWPELARLHRAGIVNAWTSNVDNPNQVNLRRMRVTGVVTTQGSAPGSAEY 175  
 QY 1204 ISQFIWYSLDGKKWQYRNGSTGLMVFFGNVDSSGKINENPPIIARYIRLHPHYS 1263  
 Db 176 MKTFKVAYSTGRKFPQIOGAESGDKIFMGNLNSGLKVNLFEPVLEVQYVRLPFIICH 235  
 QY 1264 IRSLRMLMGCDLNSCMPLGMESKASISDAQITASSYFTN--MPATWSPSKARLHLOG 1320  
 Db 236 RGCTLAPELLGCELSCGAEPGLKNDTPNKQITASSFYRTWGLSAPSWPFFARLDNQG 295  
 QY 1321 RSNAPRPQVNPKEWLOVDFOKTMKVTGVTQGVKSLTSMYKFLISSSQDGHQWTLF 1380  
 Db 296 KFNWTAQNSASEWLQIDLGSQRVGTIITQGDARDFGHIQYVAAKYAVYDDGVSWTEY 355  
 QY 1381 FQNGKV--KVFQGNDSFTVNSLDPLLTRYLIRHPQSVWVHQIALRMEVLGC 1432  
 Db 356 RDQGALEGKIFPGNLDNSHKKNMFETPLTRFVRLPVAWNRITLRVELLGC 409

## RESULT 13

MFGM\_BOVIN STANDARD; PRT; 427 AA.  
 AC Q95114; P79344; Q27959;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8)  
 DE (MGP57/53) (PAS-6/PAS-7 glycoprotein) (MFGM) (Sperm surface protein  
 DE SP47) (BP47) (Components 15/16).  
 GN MFGM8.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.

OX NCBI\_TaxID=9913;  
 RN [3]  
 RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CARBOHYDRATE-LINKAGE SITES.  
 RC STRAIN=Holstein; TISSUE=Mammary gland;  
 RX MEDLINE=97008954; PubMed=8856064;  
 RA Hvarregaard J., Andersen M.H., Berglund L., Rasmussen J.T.,  
 RA Petersen T.E.;  
 RT "Characterization of glycoprotein PAS-6/7 from membranes of bovine  
 RT milk fat globules.";  
 RL Eur. J. Biochem. 240:628-636 (1996).  
 RN [2]  
 RP SEQUENCE OF 18-427 FROM N.A.  
 RC TISSUE=Mammary gland;  
 RX MEDLINE=96125736; PubMed=8541316;  
 RA Aoki N., Kishi M., Taniguchi Y., Adachi T., Nakamura R.,  
 RA Matsuoka T.;  
 RT "Molecular cloning of glycoprotein antigens MGP57/53 recognized by  
 RT monoclonal antibodies raised against bovine milk fat globule  
 RT membrane.";  
 RL Biochim. Biophys. Acta 1245:385-391 (1995).  
 RN [3]  
 RP SEQUENCE OF 19-427 FROM N.A.  
 RC TISSUE=Testis;  
 RX Russlin M.A.;  
 RT Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE OF 140-146; 174-187; 233-246 AND 422-427.  
 RC TISSUE=Milk;  
 RX MEDLINE=93250576; PubMed=8485470;  
 RA Mather I.H., Banghart L.R., Lane W.S.;  
 RT "The major fat-globule membrane proteins, bovine components 15/16 and  
 RT guinea-pig GP 55, are homologous to MGP-E8, a murine glycoprotein  
 RT containing epidermal growth factor-like and factor V/VIII-like  
 RT sequences.";  
 RL Biochem. Mol. Biol. Int. 29:545-554 (1993).  
 CC -1- FUNCTION: PROBABLY ASSOCIATES WITH PHOSPHOLIPIDS ON THE SURFACE OF  
 CC MAMMARY EPITHELIAL CELLS AND MILK FAT GLOBULES. ZONA PELLUCIDA-  
 CC BINDING PROTEIN.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event-Alternative splicing; Named isoforms=2;  
 CC Names=Long;  
 CC IsoId=Q95114-1; Sequence=Displayed;  
 CC Names=Short;  
 CC IsoId=Q95114-2; Sequence=VSP\_001398;  
 CC -1- TISSUE SPECIFICITY: MILK AND SPERMATOZOAN  
 CC -1- PTM: THE 2 O-LINKED GLYCANS CONSTITUTE OF GAL, GLCNAC AND FUC, WITH  
 CC PROBABLY FUC AS REDUCING TERMINAL SUGAR.  
 CC -1- SIMILARITY: Contains 2 EGF-like domains.  
 CC -1- SIMILARITY: Contains 2 F5/8 type C domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X91895; CAA62997.1; -;  
 DR EMBL; S80643; AAB35894.2; -;  
 DR EMBL; Y11719; CAA72406.1; -;  
 DR PIR; S74211; S74211.  
 DR HSP; P00740; 1IXA.  
 DR InterPro; IPR006209; EGF\_like.  
 DR InterPro; IPR000421; FA58\_C.  
 DR InterPro; IPR006210; IEGF.  
 DR Pfam; PF00008; EGF; 2.  
 DR Pfam; PF00754; F5\_F8\_type\_C; 2.  
 DR SMART; SM00181; EGF\_2.  
 DR SMART; SM00231; FA58C; 2.  
 DR PROSITE; PS00022; EGF\_1; 2.  
 DR PROSITE; PS01186; EGF\_2; 2.  
 DR PROSITE; PS01285; FA58C\_1; 2.



```

DR InterPro; IPR006209; EGF like.
DR InterPro; IPR000421; F58 C.
DR InterPro; IPR006210; IEGF.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF00754; F5 F8 type C; 2.
DR PRINTS; PR00010; EGFELOOD.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00231; FAS8C; 2.
DR PROSITE; PS00022; EGF; 1.
DR PROSITE; PS01186; EGF; 2; 1.
DR PROSITE; PS01285; FAS8C; 1; 2.
DR PROSITE; PS01286; FAS8C; 2; 2.
DR PROSITE; PS00022; FAS8C; 3; 2.
DR SIGNAL; Glycoprotein; Milk; Repeat; EGF-like domain; Amyloid.
KW SIGNAL
FT CHAIN 1 23
FT CHAIN 24 387
FT CHAIN 202 387
FT CHAIN 268 317
FT CHAIN 24 67
FT DOMAIN 24 225
FT DOMAIN 70 225
FT DOMAIN 230 387
FT SITE 46 48
FT SITE 27 38
FT DISULFID 32 55
FT DISULFID 57 66
FT DISULFID 70 225
FT DISULFID 212 216
FT DISULFID 230 387
FT CARBOHYD 238 238
FT CARBOHYD 325 325
FT CARBOHYD 329 329
FT CARBOHYD 350 350
FT CARBOHYD 387 AA; 43123 MW; 2EE6571DEC83782D CRC64;
SQ SEQUENCE 387 AA; 43123 MW; 2EE6571DEC83782D CRC64;

Query Match 7.6%; Score 588; DB 1; Length 387;
Best Local Similarity 7.3%; Pred. No. 3.9e-31;
Matches 132; Conservative 69; Mismatches 125; Indels 28; Gaps 9;

QY 1093 EMILPSKAGIWRVECLIGEHLHAGMSTLFLVSNKQTPFLGMSGHIRDFQITASG----- 1147
DB 48 DVFPF-----YTCTCLKG---YAGNHC-----ETKVEPLGMENGIANSQIASSVRVTF 95
QY 1148 -QYGWAPKLARLHYSGINAW--STKEPFSWKVILLAPMIHGHTQGARQKPFSLYI 1204
DB 96 LGLQHWPELARLNAGVNAWTPSSNDNPFQIQLNRRMWTGVVTVQASRLASHEYL 155
QY 1205 SQFIIMYSLDGKKWOTYRGNSTGLMVFYFGNVDSGGIKHNIENPPIIARVIRLHPHTYSI 1264
DB 156 KAFVAYSINGHEFD-FIHVNKKHKEFGVGNKNNAVHNLFFETPVEAQVRLYPTSCHT 214
QY 1265 RSTLRMELMGCDLNSCMPLGMSKAISSDAQITASSYF----TNMFATWSPSKARLHLOQ 1320
DB 215 ACTLRPELLGCELGCANPLGLKNNSPDKQITASSYKWTGLHLF-SWNPVSARLDKQG 273
QY 1321 RSNANPQVNPKEWLOVDQKTKVTGVTQGVKSLTSMYKVEFLISSQGHQHTLF 1380
DB 274 NFNAWAGSYGNDQLQVLDLGSSEKVTGIITQGARFSGVQFVASYKVAYSNDNSANWTEY 333
QY 1381 F--QNGKVKVQFQGNDSFTPVNSLDPDLTLRLYLRHPQSVHQAIRLEMEVLG 1432
DB 334 QDPRTGSSKIFPGNWNHSHKKNLFFETPIILARYVIRLFLVAWHNRIARLELLOG 387

RESULT 15
NRP2 HUMAN
ID NRP2 HUMAN STANDARD; PRT; 931 AA.
AC O60462; O14820; O14821;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neuropilin-2 precursor (Vascular endothelial cell growth factor 165 receptor 2).
GN NRP2 OR VEGF165R2.

```

```

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS A0 AND A17).
RX MEDLINE=97470888; PubMed=9331348;
RA Chen H., Chedotal A., He Z.-G., Goodman C.S., Tessier-Lavigne M.;
RT "Neuropilin-2, a novel member of the neuropilin family, is a high
RT affinity receptor for the semaphorins Sema E and Sema IV but not Sema
RT III.";
RL Neuron 19:547-559(1997).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM A22).
RX TISSUE=Breast;
RA Soker S., Takashima S., Miao H.-Q., Neufeld G., Klagesbrun M.;
RT "Neuropilin-1 is expressed by endothelial and tumor cells as an
RT isoform-specific receptor for vascular endothelial growth factor.";
RL Cell 92:735-745(1998).
RN [3]
RP CHARACTERIZATION
RX MEDLINE=20309748; PubMed=10748121;
RA Gluzman-Poltorak Z., Cohen T., Herzog Y., Neufeld G.;
RT "Neuropilin-2 and neuropilin-1 are receptors for the 165-amino acid
RT form of vascular endothelial growth factor (VEGF) and of placenta
RT growth factor-2, but only neuropilin-2 functions as a receptor for
RT the 145-amino acid form of VEGF";
RL J. Biol. Chem. 275:18040-18045(2000).
CC -1- FUNCTION: HIGH AFFINITY RECEPTOR FOR SEMAPHORINS 3C, 3F, VEGF-165
CC AND VEGF-145 ISOFORMS OF VEGF, AND THE PLGF-2 ISOFORM OF PGF.
CC -1- SUBUNIT: NEUROPIILIN-2 PROBABLY FORMS AN HETEROMERIC COMPLEX WITH
CC NEUROPIILIN-1 IN ORDER TO BE A FUNCTIONAL SEMAPHORIN 3C RECEPTOR.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=3;
CC Name=A22;
CC IsoId=O60462-1; Sequence=Displayed;
CC Name=A0;
CC IsoId=O60462-2; Sequence=VSP_004342;
CC Name=A17;
CC IsoId=O60462-3; Sequence=VSP_004341;
CC -1- SIMILARITY: BELONGS TO THE NEUROPIILIN FAMILY.
CC -1- SIMILARITY: Contains 2 CUB domains.
CC -1- SIMILARITY: Contains 2 F5/8 type C domains.
CC -1- SIMILARITY: Contains 1 MAM domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR ENBL; AF022859; AAC51788.1; -
DR ENBL; AF022860; AAC51789.1; -
DR ENBL; AF016098; AAC12922.1; -
DR HSSP; P12259; 1CZT.
DR Genew; HGNC:8005; NRP2.
DR MIM; 602070; -
DR GO; GO:0005624; C:membrane fraction; TAS.
DR GO; GO:0004872; F:receptor activity; TAS.
DR GO; GO:0005021; F:vascular endothelial growth factor receptor. . .; TAS.
DR GO; GO:0007411; P:axon guidance; TAS.
DR InterPro; IPR000859; CUB domain.
DR InterPro; IPR000421; FAS8 C.
DR InterPro; IPR000998; MAM domain.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR Pfam; PF00629; MAM; 1.
DR PRINTS; PR00020; MAMDOMAIN.
DR SMART; SM00042; CUB; 2.

```

DR SMART; SM00231; FA58C; 2.  
DR SMART; SM00137; MAM; 1.  
DR PROSITE; PS01180; CUB; 2.  
DR PROSITE; PS01285; FA58C.1; 2.  
DR PROSITE; PS01286; FA58C.2; 2.  
DR PROSITE; PS01287; FA58C.3; 2.  
DR PROSITE; PS00222; FA58C.3; 2.  
DR PROSITE; PS00660; MAM 2; 1.  
KW Transmembrane; Glycoprotein; Neurone; Signal; Repeat; Receptor;  
KW Alternative splicing.  
FT SIGNAL 1 20 OR 22 (POTENTIAL).  
FT CHAIN 21 931 NEUROFILIN-2.  
FT DOMAIN 21 864 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 865 889 POTENTIAL.  
FT DOMAIN 890 931 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 28 142 CUB 1.  
FT DOMAIN 149 267 CUB 2.  
FT DOMAIN 277 427 FS/8 TYPE C 1.  
FT DOMAIN 434 592 FS/8 TYPE C 2.  
FT DOMAIN 642 802 MAM.  
FT DOMAIN 671 674 POLY-SER.  
FT DISULFID 28 55 BY SIMILARITY.  
FT DISULFID 83 105 BY SIMILARITY.  
FT DISULFID 149 175 BY SIMILARITY.  
FT DISULFID 208 230 BY SIMILARITY.  
FT DISULFID 277 427 BY SIMILARITY.  
FT DISULFID 434 592 BY SIMILARITY.  
FT CARBOHYD 152 152 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 157 157 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 629 629 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 839 839 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT VARSPLIC 809 813 Missing (in isoform A17).  
FT VARSPLIC 809 830 Missing (in isoform A0).  
FT CONFLICT 602 602 /FTID=VSP 004342.  
FT SEQUENCE 931 AA; 104830 MW; 270CBAB69A0A797C CRC64;  
Query Match 6.1%; Score 469.5; DB 1; Length 931;  
Best Local Similarity 28.5%; Pred No. 8.1e-23;  
Matches 166; Conservative 87; Mismatches 185; Indels 145; Gaps 32;  
QY 934 DEFDC-AWAYPSVDLEK-----DVHSLGILGLLVCHTNL-----NPAHGRQVTVQBEFA 983  
DB 79 EKHDCKYDFEIRDDGDSADLLGKHCNIAPTTIISGSMIYIKFTSDYARQGA--GFS 136  
QY 984 LFFETFDKSWYFTENMERNCRAPNIQMEDPTKENYRPHAINGYIMDTLPGLVMAQD 1043  
DB 137 LRYELF-KTGS-----EDCSKNFTSP-NGTIESPGFPEKYP----- 170  
QY 1044 QRIRWYLLSMGNSNIHSHFSGHVFTVRKKEVKALYNLYPGVFTVEMLPKAG--- 1100  
DB 171 -----HNLDCFTILAKPKMEIILQFL---IFD-LEHDPLOVGECD 207  
QY 1101 -----IW-----RVECLIGH-----LHAGMT-----LFLV 1122  
DB 208 CKYDMLDIWDGIPHYGLPGIKYCGKTPSELRSSTGILSTPHTDMAVAKDGFSAARYLV 267  
QY 1123 YSN-----KCQTPGLMASGHIEDFOITASQY--QWAPKLARLHYSGSINAW-----ST 1170  
DB 268 HOEPLNFQCNVPLGMSGRIANEQISASTYSVSGRWTPQOSRLH--GDDNGWTPNLDN 325  
QY 1171 KEPFSWIKVDLAPMIIHGIKTQGA--ROKFSYLSIOFIIMYSLDGKKWQTYRGNSTGT 1228  
DB 326 KE---YLQVDLFLTLTAIATQGAISRETQNGYTVKSKLEVSTNGEDMNVYRHGKH- 381  
QY 1229 LMVFGNVDSGKIHNIFFPIIARYIRLHPHYISIRTLRMELMGCDLNS--CSMPLGM 1286  
DB 382 -KVFQANNDATVNLKHLAPLITFEVIRPQVTHSGIALRLLEFGCRVTDAPCSNMLGM 440  
QY 1287 ESKAISDAQITASSYFTNMFATWPSKARLHQGSNAW---RPQVNNPKWLQVDFOKT 1343  
DB 441 LSGLIADSQISASS--TOEY-LWSPSAARL-VSSRS-GWFFRIPQAPQGEWLQVDLQTP 495

QY 1344 MKVTGVTTQG-----VKSLTSMVYKFEFLISSODGHQWTLFFONGKV---KVFQGNOD 1394  
DB 496 KTVKGVIIQAGRGDSITAVEARAFVRKFKVSYSLNGKDWZ--YIQQPRTQPKLPEGNMH 554  
QY 1395 SFTPVVNSLDPPLLTRYLRHHPQSWVHQ-IALRMEVLGCEAQD 1436  
DB 555 YDTPDIRRFD-PIPAQYVRYVPERWSPAGIGRLVGLGCDWTD 596

Search completed: December 9, 2003, 16:54:09  
Job time : 62 secs



Result No.	Query			DB	ID	Description
	Score	Match	%			
1	6272	81.5	2343	6	O18806	O18806 canis famil
2	6231	81.0	2343	6	O62730	O62730 canis famil
3	2486.5	32.3	2119	13	Q90X47	Q90X47 brachydanio
4	2386.5	31.0	2224	4	O43737	O43737 homo sapien
5	2369	30.8	2183	11	O88783	O88783 mus musculu
6	1799	23.4	1157	11	Q92024	Q92024 mus musculu
7	1784	23.2	1157	11	Q92088	Q92088 rattus norv
8	1747	22.7	1158	4	Q9BQS7	Q9BQS7 homo sapien
9	1742	22.6	1158	4	Q9C058	Q9C058 homo sapien
10	1710	22.2	1104	4	O75180	O75180 homo sapien
11	1709.5	22.2	1084	11	Q9J197	Q9J197 rattus norv
12	1618	21.0	1048	6	Q8XT27	Q8XT27 ovis aries
13	1534.5	20.0	355	11	O8BQ43	O8BQ43 mus musculu
14	1285	16.4	847	11	O8C4S2	O8C4S2 mus musculu
15	1221	15.9	782	4	Q75659	Q75659 homo sapien
16	1104	14.4	216	4	Q14286	Q14286 homo sapien

QY 1 ATRRYLGAVELSDYMSD-LGELPVDARFPFPRVPKSPFNTSVVYKKTIFVEFTVHIF 59

Db	20	ATRKYLGAVELSWYQMSDLLSALHADTSFSSRVPGSLPLTTISVTYKTVFVFETDDLF	79
Qy	60	NIAPRPWMLGPTTQAEVYDTVTITLKMASHPVSLHAGVSYWKASGAEYDDQTS	119
Db	80	NIAPRPWMLGPTTQAEVYDTVTITLKMASHPVSLHAGVSYWKASGAEYDDQTS	139
Qy	120	QREKDDKVPFGSGHTVVOVLKENGPMASDPLCLTVSYLSHVDLVKDLNGLIGALLVC	179
Db	140	QKKEEDNVIPGSHYTVVOVLKENGPMASDPLCLTVSYLSHVDLVKDLNGLIGALLVC	199
Qy	180	REGSLAKEKQTTLHKFTLLFAVDEGKSHSETKNSLMQDRDAASARAWPKMHTVANGYN	239
Db	200	KEGSLAKERTOTLQEVLLFAVDEGKSHSETNASLTQ-----AEAQHELHTINGYN	253
Qy	240	RSLEPLIGCHRSYVHVIGMGTTPVHSIFLGHTFLVRNHRQASLEISPTIFLTAOTL	299
Db	254	RSLEPLTVCHRSYVHVIGMGTTPVHSIFLGHTFLVRNHRQASLEISPTIFLTAOTL	313
Qy	300	LMDLGOFLFCHISSHOHGMBAVVKVDSCEPEPQLRMKNREAEYDDDLTDSMDVVR	359
Db	314	LMDLGOFLFCHIPSHOHGMBAVVKVDSCEPEPQLRMKNREAEYDDDLTDSMDVVR	372
Qy	360	FDDNSPFSQIRSVAKGPKTWVHYIAAEEDWDYAPLVAPDSDRSYKSYQLNNGPQRI	419
Db	373	FDDSSSPFIQIRSVAKGPKTWVHYIAAEEDWDYAPSGPTENDRSHKNLYLNGPQRI	432
Qy	420	GRKYKVRPMAYTDETFKTRATOHESGILGPLLYGEVGDTLIIIFKNQASRPYNTYPHG	479
Db	433	GRKYKVRPMAYTDETFKTRATOHESGILGPLLYGEVGDTLIIIFKNQASRPYNTYPHG	492
Qy	480	ITDVRPLYSRRLPKGVKHLKDPILPGEIFPKYKWTVTVEDGPKSDPRCLTRYSSFVNM	539
Db	493	INVTFLHTGRLLPKGVKHLKDPILPGEIFPKYKWTVTVEDGPKSDPRCLTRYSSFVNM	552
Qy	540	ERDLASGLIGPLLI CYKESVDQRGNQIMSDKRNVLFSVFDENRSWYLTENIORFLPNA	599
Db	553	ERDLASGLIGPLLI CYKESVDQRGNQIMSDKRNVLFSVFDENRSWYLTENIORFLPNA	612
Qy	600	GVQLEPPEQASIMHSINGYVDSLSQLSVCLHEVAYWYLSIGATDLSVFFSGYTFK	659
Db	613	VQPHDPPEQLSIMHSINGYVDSLSQLSVCLHEVAYWYLSIGATDLSVFFSGYTFK	672
Qy	660	HKVVYEDTLTLFPFSGETVFMSENPLATLGCHNSDFRNGMTALLKYSSCDKNTGDYV	719
Db	673	HKVVYEDTLTLFPFSGETVFMSENPLATLGCHNSDFRNGMTALLKYSSCDKNTGDYV	732
Qy	720	EDSYEDISAYLLSKNNAIBRSP-----	742
Db	733	EDTYEDIPTLNENNVIKPRFSQNSRHPSTKEKOLKATTPENDIEKIDILQSGERTOL	792
Qy	743	-----	742
Db	793	IKAQSVSSDLLMLLQNPTRGLFISDLREATDRADHSGAIERNKGPEVASIRPEL	852
Qy	743	-----	742
Db	853	RHSEDEFTPEBELQLRINENGTNTVELKKLDLKISSSDSLMSTPTIPSDKLAATE	912
Qy	743	-----	742
Db	913	KTGSLGPPNWSVHFNHSLGTIVFGNNSSHLIQSGVPLELSEEDNDSKLLLEAPLMNIQESS	972
Qy	743	-----	742
Db	973	LRENVLMSNRLFKBERIRGPASLJKONALFKVNISSVKTNRAPVNLTTNRKTRVAPT	1032
Qy	743	-----	742
Db	1033	LIENSTSVQDMLERNTEPFKEVTSLIHNFTMDRNTTALGLNHSNKTTLISKVNEVAH	1092
Qy	743	-----	742
Db	1093	QKEDVPVLAENPDLSKIPFLPDWIKTHGKNSLSSEQRSPKQTLGSEKSKVQON	1152
Qy	743	-----	742
Db	1153	FLSEKVVVGEDEFTKOTELQEBIFPNKNSIFFANLANVQENDTYNOEKSPBEIERKXL	1212
Qy	743	-----	742
Db	1213	TQENVALPOAHMTI GTKNFLKNLFLLLSTKONVAGLEEQYPTPILOTRSLNDSPHSEGIH	1272
Qy	743	-----	742
Db	1273	MANFSKIREANLEGLGNQTNQVREFPSTTRMSSNASQHVITQRCRSIKQPRLSQGEI	1332
Qy	743	-----	742
Db	1333	KPERKVIANDTSTQMSKNMNYLAQGLTQI EYNEKEKRAITQSPISDCSMRNVHTIQMND	1392
Qy	743	-----	742
Db	1393	SALPVAKSASPSVRHTDUTKIPSHNSHLPASACNTYFRERTSGVBQSHFLQEAEN	1452
Qy	743	-----	742
Db	1453	NLSLAFVTILGITEGQKFSGLCKSATNQPMYKXLENTVLLQPLGSETSKVELLSQVHVD	1512
Qy	743	-----	742
Db	1513	QEDSPPTKTSNDSPGHLDLMGKIFLQKTOGPVKMNKNSPGKVPFLKWATESSEKIPSKL	1572
Qy	743	-----	742
Db	1573	LGVLAWDNHYDQIIPSEBWKSKSQTNTAFKPKDTIPLGPCENNDSTAAINEGQDKPQ	1632
Qy	743	-----SONPPVLKRHQREITRTTLOSDBEIDYDDTISVEMKEDFDIYD	787
Db	1633	REAWAKQEGEPGLCSQNPVSKHQREITVTLQPEEDKFYDDTFSIEMKREBFDIYD	1692
Qy	788	EDENSGPSFKKTRHYFIAAVERLWDYGMSSSPHVLNRNAQSGVQPKVVFQFETDG	847
Db	1693	DYENQGLARSFQKTRHYFIAAVERLWDYGMSSSPHVLNRNAQSGVQPKVVFQFETDG	1752
Qy	848	SFTQPLRGELNEHGLIGPYIRAEVEDNIMVTFNQAASRPYSFSSLSIYSEEDQOQAE	907
Db	1753	SFTQPLRGELNEHGLIGPYIRAEVEDNIMVTFNQAASRPYSFSSLSIYSEEDQOQAE	1812
Qy	908	PRKNFVKNETKTYFWKQVHMAPTKDFCDKAWAYFSDVDLEKDVHSLGLIGLIVCHTN	967
Db	1813	PRKFNVPNETKIYFWKQVHMAPTKDFCDKAWAYFSDVDLEKDVHSLGLIGLIVCHTN	1872
Qy	968	TLPNAGRQVTVQEPFALFTTIPDETYSWYFTENMERNCRAPCNIQMEDPTPKENYRFAH	1027
Db	1873	TLPNAGRQVTVQEPFALFTTIPDETYSWYFTENMERNCRAPCNIQMEDPTPKENYRFAH	1932
Qy	1028	NGYIMDTLPLGVMAODQIRIWRVLLSMGNSNIHSTHESGHVPTVKKKBYKXALVNLVPG	1087
Db	1933	NGYIMDTLPLGVMAODQIRIWRVLLSMGNSNIHSTHESGHVPTVKKKBYKXALVNLVPG	1992
Qy	1088	VFETVEMLPFSKAGIWRVLECLIGEHLHAGMSTLFLVYSNKCOTPLGMASGHIRDFOITAG	1147
Db	1993	VFETVEMLPFSKAGIWRVLECLIGEHLHAGMSTLFLVYSNKCOTPLGMASGHIRDFOITAG	2052
Qy	1148	QYGQWAPKARLHYSGSINAWSTKPEFWSIWKVLLAPMIHGIKTQAGQKPFSSLIYSQF	1207
Db	2053	QYGQWAPKARLHYSGSINAWSTKPEFWSIWKVLLAPMIHGIKTQAGQKPFSSLIYSQF	2112
Qy	1208	LIYSLDCKKQVTVRGSTGTLMVFFGNVDSGKIKHNI FNPIIARYIRLPHTHYSIRST	1267
Db	2113	LIYSLDCKKQVTVRGSTGTLMVFFGNVDSGKIKHNI FNPIIARYIRLPHTHYSIRST	2172
Qy	1268	LRMELMGCDLNSCSNPLGMESKALSDAQITASSYFTNNFATWSPSKARLHLOGRNWRP	1327
Db	2173	LRMELMGCDLNSCSNPLGMESKALSDAQITASSYFTNNFATWSPSKARLHLOGRNWRP	2232

```

QY 1328 QVNPKEVLOYDFQKTKMKTGTTQGVKSLLTSMVTKBFLISSQDGHQWTLFFPONGKVK 1387
Db 2233 QANNPKEVLOYDFRKTMTGTTTQGVKSLLTSMVTKBFLISSQDGHQWTLFFLQNGKVK 2292

QY 1388 VFQGNQDSFTPVNSLDPLLRXYLRHQPQSVHQAIALRMEVLGCEAQ 1435
Db 2293 VFQGNRDSSTPVNRLEPPLVARYLRLPQSWAHHAIRLEVLGCDTQ 2340

RESULT 2
O62730 PRELIMINARY; PRT; 2343 AA.
AC O62730;
DT 01-AUG-1998 (TRENBLrel. 07, Created)
DT 01-AUG-1998 (TRENBLrel. 07, Last sequence update)
DE 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DB Factor VIII.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Kidney, and Spleen;
RA Gordy P.W., Bowen R.A.;
RT "Characterization of the canine factor VIII cDNA.";
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
DR EMBL; AF049489; AAC05384.1; --
DR HSSP; P00451; 1CFG
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR000421; FA58_C.
DR Pfam; PF00394; Cu-oxidase; 3.
DR Pfam; PF00754; F5 F8 type C; 2.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
DR PROSITE; PS00079; MULTICOPPER OXIDASE1; 3.
SQ SEQUENCE 2343 AA; 265613 MW; F612D744ADAADD99 CRC64;

Query Match 81.0%; Score 6231; DB 6; Length 2343;
Best Local Similarity 53.2%; Pred. No. 0;
Matches 1238; Conservative 85; Mismatches 105; Indels 900; Gaps 4;

QY 1 ATRRYLGAVELSDYMQSD-LGELPVDAREPPRPVPSFPENTSVYKTLFVFTVHLF 59
Db 20 ATRKYLGAVELSDYMQSDLLSALHADTSFSSRVPGSLPTTSVYKTVFVFTDGLF 79

QY 60 NIAKPRPPMGLLGTQAEVYDVTIVILKNASHPVSLHAGVSVYKASGAEYDDOTS 119
Db 80 NIAKPRPPMGLLGTQAEVYDVTIVILKNASHPVSLHAGVSVYKASGAEYDDOTS 139

QY 120 QREKDDKVPFGSHYVYVQVLEKNGPMASDPLCLTYSLSHVDLVKDLNSGLIGALLVC 179
Db 140 QREKDDNVIPGSHYVYVQVLEKNGPMASDPLCLTYSYFSHVDLVKDLNSGLIGALLVC 199

QY 180 REGSLAKEKTLTKHFTILLFAVFDGKSWHSETKNLSMQDRDAASARAWPKHHTVNGYVN 239
Db 200 KEGSLAKERQTLQEFVLLFAVFDGKSWHSETWASLTQ-----AEQHELHTINGYVN 253

QY 240 RSLPGLGCHRSKYVHYVIGMTTPEVHSIFLEGHTFLVRNHRQASLEISPTFLTAQTL 299
Db 254 RSLPGLVCHRSKYVHYVIGMTTPEVHSIFLEGHTFLVGNHRQASLEISPTFLTAQTF 313

QY 300 LMDLQGFLLFCHISSHQDGHAEYKVDSCPEEPQLRMKNNEAEYDDDLTDSMDVVR 359
Db 314 LMDLQGFLLFCHISPHQDGHAEYKVDSCPEEPQLRMKNNEED-KYDDGLGYSDMDVVS 372

QY 360 FDDNDSPSFTQIRSAVKHPTWVHYIAAEEEDNDYAPLVLPDDRYSVKQYLNNGPQRI 419
Db 373 FDDSSSPPFTQIRSAVKHPTWVHYIAAEEEDNDYAPSGPTPNDRSHKNLYLNNGPQRI 432

QY 420 GRKYKVRFMAYTDTETFKTRAIQHESGILGLPLLYGEVGDTLIIIFKNQASRPYNIYPHG 479

```

```

Db 433 GKTKYKVRFMAYTDTETFKTRAIQHESGILGLPLLYGEVGDTLIIIFKNQASRPYNIYPHG 492
QY 480 ITDVRPLYSRELPGVGHKLDFFILPGHEIFKYKWTVTVEDGPTKSDPRCLTRYSSVNM 539
Db 493 INYVPLHTGLRPGVGHKLDMPILPGHEIFKYKWTVTVEDGPTKSDPRCLTRYSSVNL 552
QY 540 ERDLASGLIGPLLI CYKESVDQGNQIMSDKRNILFSVFDENRSWYLTENIQRFPLNPA 599
Db 553 ERDLASGLIGPLLI CYKESVDQGNQIMSDKRNILFSVFDENRSWYLTENIQRFPLNPA 612
QY 600 GVQLEDEPFOASNIHMSINGVYVDSLOLSVCLHVAWYILSIGAOTDFLSVFSGYTFK 659
Db 613 VVQPHDPEFQLSNIHMSINGVYVDFNLQLSVCLHEVAWYILSVAQTDFLSVFSGYTFK 672
QY 660 HKMVEYDTLTLFPFSGTVMFMSNPGIWLIGCHNSDFRNRGNTALLKVSCKDNTGY 719
Db 673 HKMVEYDTLTLFPFSGTVMFMSNPGIWLIGCHNSDFRNRGNTALLKVSCKRNDIDY 732
QY 720 EDSYEDISAYLLSKNNAIEPRSF----- 742
Db 733 EDYTEDIPTPLANNENVIKPRSFQNSRHPSTKEKQLKATTPPENDIEKIDLQSGERTQL 792
QY 743 ----- 742
Db 793 IKAQSVSSDILLMLLQGNPTPRGLFGLDLREATDRADHSGAITERNKGPPVASLRPEL 852
QY 743 ----- 742
Db 853 RHSEDRFTPEPELQLRNLNENLGNNTTVVELKKDLKISSSDSLMTSPTIPSDKLAATE 912
QY 743 ----- 742
Db 913 KTGSLGPPNMSVHFNGHLGTVFCGNSSHLIQSGVPLELSEEDNDKLEAPLNIQESS 972
QY 743 ----- 742
Db 973 LRENVLSMESNRLFKEERIRGPASLIKONALFKVNISSVKTRAPVNLTTNRKTRVAIPT 1032
QY 743 ----- 742
Db 1033 LLIENTSVWQDIMLERNTEFKEVTSLIHNFTMDRNTTALGLNHVSNKTTLSKNVEMAH 1092
QY 743 ----- 742
Db 1093 QKKEDPVPLAENPDLSSKIPFLPDWKTHGKNLSSEORPPSPKQLTSLGSEKSVKQDN 1152
QY 743 ----- 742
Db 1153 FLGEEKVVGEDBFTKDTLQEI FPNKNSIFFANLANVQENDTYNQEKKSLEERKEKL 1212
QY 743 ----- 742
Db 1213 TOENVALPOAHWTIGTKNFKNLFLSLTKQNVAGLEEQYPTILQDTRSLNDSPHSGIH 1272
QY 743 ----- 742
Db 1273 MANFSKIREANLEGLNQTNQWVERPPSTRSSNASQHVITQGRKRSUKQPRLSQGEI 1332
QY 743 ----- 742
Db 1333 KPERKVIANDTSTQWSKNMNYLAQGTLTQIEYNEKEKRAITQSPSLDCSMRNVHTIQMD 1392
QY 743 ----- 742
Db 1393 SALFVAKESASPSVRHTDLTKIPSOHNSHLPASACNYTPRERTSGVQEGSHFLQAKRN 1452
QY 743 ----- 742
Db 1453 NLSLAFVTLGITGQOKFSLGSKATNQPMYKLENTVLLQPLGSETSKVELLSQHVHD 1512
QY 743 ----- 742

```

Db 1513 QEDSPPTKTSNDSPGHLDMGKI FLQKTQGPVKMNKTSNPGKVPFLKWATESSEKIPSKL 1572  
Qy 743 ----- 742  
Db 1573 LGVLAWNDHYDTQIPSEWKSQKSTNTAFKRXDTILPLGPCENNDSAAINEGQDRKP 1632  
Qy 743 -----SQNPVLKRHORETRITTLQSDQBEIDYDDTISVEMKEDFDYD 787  
Db 1633 REAMWAKQEPGRLCSQNPVSKHOREITVTTLQPEEDKFEYDDTTFIEMKREDFDYD 1692  
Qy 788 EDENSGPSRQKKTTHRYFIAAVERLDWYGMSSPHVLNRNRAQSGVPOFKKVVQEFPTD 847  
Db 1693 DYEDQGLRSFKKTRHYFIAAVERLDWYGMSSPHVLNRNRAQSGVPOFKKVVQEFPTD 1752  
Qy 848 SFTQPLRGELNEHLGLGPGYRAEVEDNIMWTFNQAASRPSPYSSLSISYEDDQGA 907  
Db 1753 SFTQPLRGELNEHLGLGPGYRAEVEDNIMWTFNQAASRPSPYSSLSISYEDDQGA 1812  
Qy 908 PRKNVKNETKTYFWKVOHMAPTKDEPDKAMAYFSDVDLEKDVHSLGLPLLVCHTN 967  
Db 1813 PRKFWNETKTYFWKVOHMAPTKDEPDKAMAYFSDVDLEKDVHSLGLPLLVCHTN 1872  
Qy 968 TLNPAHQVQVQEPALFETIPDETYSWYFTENMEERNCRAPCNIQMEDPTFKENYRFAI 1027  
Db 1873 TLNPAHQVQVQEPALFETIPDETYSWYFTENMEERNCRAPCNIQMEDPTFKENYRFAI 1932  
Qy 1028 NGYINDTLPGLVMAODQIRWYLLSMGNSNENIHSIFSGHVTFRKBEYKXALNLYPG 1087  
Db 1933 NGYINDTLPGLVMAODQIRWYLLSMGNSNENIHSIFSGHVTFRKBEYKXALNLYPG 1992  
Qy 1088 VPETVEMLPSPKAGIWRVCLIGEHLHAGMSTLFLVYSNKCOTPLGWSGHIRDFOITASG 1147  
Db 1993 VPETVEMLPSPKAGIWRVCLIGEHLHAGMSTLFLVYSNKCOTPLGWSGHIRDFOITASG 2052  
Qy 1148 QYGWAPKARLHYSGSINAMSTKPFWSKWKVDLLAPMIHIGIKTQAGRKPFSSLYISQF 1207  
Db 2053 QYGWAPKARLHYSGSINAMSTKPFWSKWKVDLLAPMIHIGIKTQAGRKPFSSLYISQF 2112  
Qy 1208 IIMYSLDGKKWQYEGNSTGTLWFFGNVDSGIGKHNIFNPILARIYLRHPTHYSIRST 1267  
Db 2113 IIMYSLDGKKWQYEGNSTGTLWFFGNVDSGIGKHNIFNPILARIYLRHPTHYSIRST 2172  
Qy 1268 LRMELMGCDLNSCMPLGWSKAISDAQITASSYFTNPFATWSPSKARLHLQGRNWRP 1327  
Db 2173 LRMELMGCDLNSCMPLGWSKAISDAQITASSYFTNPFATWSPSKARLHLQGRNWRP 2232  
Qy 1328 QVNNPKWLQVDFQTKMKTGVTGTQGVKSLLTSMYVKGFLISSQDGHOWTLFFQNGKVK 1387  
Db 2233 QVNNPKWLQVDFQTKMKTGVTGTQGVKSLLTSMYVKGFLISSQDGHOWTLFFQNGKVK 2292  
Qy 1388 VFQGNQDSFTPVNSLDPLLTLYLRHQPQSVHQAIALRMEVLGCEAQ 1435  
Db 2293 VFQGNQDSFTPVNSLDPLLTLYLRHQPQSVHQAIALRMEVLGCEAQ 2340

## RESULT 3

Q90X47 PRELIMINARY; PRT; 2119 AA.  
AC Q90X47;  
DT 01-DEC-2001 (TremBLrel. 19, Created)  
DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)  
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)  
DE SC:hz2015.3 (Novel protein similar to vertebrate coagulation factor V and VII).  
GN SC:hz2015.3.  
OS Brachydanio rerio (Zebrafish) (Danio rerio).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
OC Cyprinidae; Danio.  
OX NCBI\_taxid=7955;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Lloyd D.;

RL Submitted (OCT-2001) to the EMBL/GenBank/DBAJ databases.  
CC -L- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.  
DR EMBL; ALS90146; CAC94896.1; -.  
DR InterPro; IPR001117; Cu-oxidase.  
DR InterPro; IPR000421; FAS8\_C.  
DR Pfam; PF00394; Cu-oxidase; 2.  
DR Pfam; PF00754; F5\_F8\_type\_C; 2.  
DR SMART; SM00231; FAS8C\_2.  
DR PROSITE; PS01285; FAS8C\_1; 2.  
DR PROSITE; PS01286; FAS8C\_2; 1.  
DR PROSITE; PS00079; MULTICOPPER\_OXIDASE1; 3.  
SQ SEQUENCE 2119 AA; 240643 MW; DC0806FFA8761E6 CRC64;

## Query Match

Best Local Similarity 32.3%; Score 2486.5; DB 13; Length 2119;  
Matches 600; Conservative 262; Mismatches 491; Indels 815; Gaps 36;

Qy 3 RRYTLGAVELSWDQMQLGELPVDARPPRPVKSPFPNTSVVYKTLTFVFTVHLFVIA 62  
Db 28 RHYIAAVNINWDYTSQQ-----RTGQSYKVVYREYN-EGFKQP 67  
Qy 63 KPRPFWMLLQPTTQAEVYDVTWITLKNMASHPVSLHAGVSYWKASEGAYDDQTSORE 122  
Db 68 KAHLSSGLLQPTLRGQEGDTIIVFRNADHPCLPHGLAYGKQSGSLYFDNTSLLE 127  
Qy 123 KEDDKVPFGSGHTYVQVLKENGPMASDPLCLTYSLYSHVDLVKDLNSGLIGALLVCREG 182  
Db 128 KNDVIOQGEHTYQWDTSDVTPTAADPPCITYSLSHFDIVRDYNTGLGPMILICKG 187  
Qy 183 SLAKEKTQTLH---KFILLFAVDFEGKSWHSETKNSLMQDRDAASARAPKMHVTWGVYN 239  
Db 188 TLDDSGNQ-IHFQESVLLFGVFDENKSWY-----TGDSPPQLVNKYTINGYN 236  
Qy 240 RSLPGLTCHRSKSYVHVIGMTTPEVHSIFLEGHTFLVRNHRQASLEISITFLTAQTL 299  
Db 237 GSPDLDICAHSKYKSWHLGLGMSSELPFSVHFNQGLVLDHGKTSAVGIISGTATASMT 296  
Qy 300 LMDLQGLLFCFCHSHQHDGMEAVYKVDSCPE--EPQLRMKNNEAEYDDDLTSDMDV 357  
Db 297 GVHGRMLVSSHISKHLAEGLYLNIRKDCDEYAPKRLRTIEQ----- 340  
Qy 358 VRFDNDSPSFIQIRSAKHKPTWVHYIAAEEDWDYAPLVLPADDRSYKSOVLNNGQP 417  
Db 341 -----KKESQWYMAAEVWDYAPNPNMDGDFSKYLKQGP 382  
Qy 418 RIGKYKVRPMAYTDEFTTR---BAIQHESGILGPLLYGEVGTLLIIFPNQASRPYN 474  
Db 383 RIGKYKAVPTQYKDGMFKEARDKQRELGLGPVIRAIIRDIKIIVFNKASRPYS 442  
Qy 475 IYPHGITDVRPLYSRRLPKGVKHLKDFPLPGLPFIKPKYKWTYVDEGPKSPCLTRYYS 534  
Db 443 IYPHGLTIDKAAEGASYPQGN--QTSYVQGEITYTWTSTEDVPTSDPRLCTRMVH 500  
Qy 535 SFVNMERDLASGLIGPLLI CYKESVDQKGNQIMSDKRNVLFSVFDENRSHYLTENIORF 594  
Db 501 SAVDAPRDIASGLVGPLLI CKSQSLNKNVQLKADKEQHAFVTFDENKSHYQENINTY 560  
Qy 595 LPNPAGVQLEDPEFQASNMHSINGYVFDLSIQ-LSVCLHEVAYWYILSIGAQTFLSVFF 653  
Db 561 CSPDKVKKDDPEFYKSNVMTINGYVYESQELGFGCHGEIVTWHVSSVGQDYIQTATP 620  
Qy 654 SGYTFKHMYVDLTLPFPFSGETVFMENPGMLWILGCHNSDFRNGMTALLKVSCK 713  
Db 621 YGHTFELNREEDILSLFPMGTGETITWNVNIGIWLASLNSHSDTKGMVKFKDLCEFR 680  
Qy 714 NTGDY-----YE-----DSYEDISAYLL----- 731  
Db 681 ---DYVLEYDEGKFTAWKPTTINEIKKEPVAPRDVDVDEYSDLPASTLNLTFFNVK 737  
Qy 732 -----SK 733  
Db 738 DEVEIIDLTFLDQDDGLLPVIEKSLGSSNENLNHATLQSFIEHTHGLMBEGDLKDESS 797

```
QY 734 NNAI-----EPRFS 743
Db 798 NKVLNDSTDKALLETTTFDSNRVVVALNETDSIILDFPIVERKVSAPKMEPESVT 857
QY 744 QN-----PVLKRHQREIIRT-----TLQSD--- 764
Db 858 MNFKTEHINSLSERINAIYSPITETNINWTEHTDFTSITPFGSTGEMNFTLEDDTAL 917
QY 765 -----QEEI-----DYD-----DTISVENK- 779
Db 918 LNSSESEPLQNSNSENRIAFQELNNAKGDVDVSNNSVKQIKPKYNYVPSGDTLNSSKI 977
QY 780 ---KEDFDIYD----- 787
Db 978 QVEEDFVLLDSSYFSEMSTTMEYDVSQDXTVKGESKETAQOGLSSTKTKYSGEILLES 1037
QY 788 ----- 787
Db 1038 PDITSAFNLSSVLNRNLSNLSNETLPMSSNATFSDSTNATSDSSTATPADFSNT 1097
QY 788 ----- 787
Db 1098 TFSNATPFDGFRISQMSDSSNATLSDSSNATLSDSSNATLSDSSNATLSDSS 1157
QY 788 ----- 787
Db 1158 SNATPDSNKTFSNATLFGVSYSSNTTSLDPSLESEMTEYLSSANDTIKSHSEV 1217
QY 788 -----EDENQ----- 792
Db 1218 SNTQLSSSESTENISLLYGSLNASSMKNDSESEEVVILNKNHSEAILTSHLDQKE 1277
QY 793 ----- 792
Db 1278 EHWGESKHEIVHKELPDHMKYVKDKSAANSKPKLEKKKYQVRKPKKGGMKTKK 1337
QY 793 -----SPRSP----- 797
Db 1338 SKDYKQPRSSFSPRGFGPSVLTTPRGSRPVSSSEDELTEKPIVIGVPRRDFNDYELIPKQ 1397
QY 798 -----OKKTHYFIAAVER 811
Db 1398 DQEADFGLLDHPBEYVEYKDYPSKTDVQALDATSQHLLKMGADKNTKTYISVEE 1457
QY 812 LWDYGMSSPHVLNRAQSGSVPOFKVVOEFTDGSFTQPLYRGELNEHLGLLGPVIRA 871
Db 1458 ENDYA-CYQORLDTAQNRPPTVRKVFYRLDSTFSDIRGEMDEHLGLLGPLIKA 1516
QY 872 EVDNIMVTPRNOASRPSYSSLIY-----BEDORQABPRKFKVKNETKTYFW 923
Db 1517 EVDQTMVFFNRASRPSYSLHANGVKYLKQMEGLSYDDESPYWKQDDAVPPNGTFTYMW 1576
QY 924 KVQHMAPTKDEPCKAWAFSDVDLEKDVHSGILGLLCHNTNLTN--PAHGRQVTVQE 981
Db 1577 TIIPKSGQNNESDCRTWTIYSAVNPEDINSGLIGLPLLCRKGTLCKPEDRR-----E 1631
QY 982 FALFFITFDTKSYFTENMERNCRAPCNQMEDPTFKENYRFAINGYIMDTLPLGLVMA 1041
Db 1632 FULLFTFDENKSWLYEENRQRIERKNRRVVM-DPNEQDNLKDPAING-IYSLKGLRMY 1689
QY 1042 QDQIRWILSMGSENNIHSIHFSGHVFTVRKKEEYKALYNLYPGVFETVEMLPKXAGI 1101
Db 1690 TNQLAKWHLINMGSPKDLHSVHFHGTQFINKELKDRHQGVVPLLPGGFATLMLPSKPL 1749
QY 1102 WRVECLLGEHLHAGMSTLF-----LVYSNCKQTPFGMASGHIRFOI 1143
Db 1750 WQLESVEGLSQORGMOTLFLFDESYYKRYVCQVLRSLYLIYFAVCDHPLGISGTVDQEQI 1809
QY 1144 TASQYQWAPKLARLHYSGINAWST-KEPFSWKVDLLAPMLIIGHIKTQGRQKPSL 1202
Db 1810 TASDTRQWTPHLARLHNTKYNWSTSEPGQLQVDFORPVVISKATQGAQKQFLTHN 1869
QY 1203 YISQFIIMYSLDGKKWQYRGNSTGT-----LMVFF-----GNVDSGKIKHNFNP 1248
```

## RESULT 4

```
O43737
ID O43737 PRELIMINARY; PRT; 2224 AA.
AC O43737;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Factor V.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RA Bird C.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
DR EMBL; Z99572; CAB16748.1; -.
DR HSP; P12259; LC2T.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR000421; FA58_C.
DR Pfam; PF00394; Cu-oxidase; 3.
DR Pfam; PF00754; F5_P8 type C; 2.
DR SMART; SM00231; FA58C_2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
DR PROSITE; PS00079; MULTICOPPER OXIDASE1; 2.
SQ SEQUENCE 2224 AA; 251673 MW; 1013108D49690EAB CRC64;
```

Query Match 31.0%; Score 2386.5; DB 4; Length 2224;  
Best Local Similarity 26.0%; Pred. No. 2.2e-169;  
Matches 592; Conservative 278; Mismatches 483; Indels 923; Gaps 35;

```
QY 3 RRYLGAVELSDWYMQSDLGELFVDARFPFRPKSPFNTSVV-YKKTLPVFETVHLFNI 61
Db 32 RQFVAAQGISWRPE-----PNSLSLSVTSFKIVREVEPY-FKK 75
QY 62 AKRPPPMGLLGTIOAEVYDVTITLKNWASHPVSLHVGVSFWKASGEAYDDQTSQR 121
Db 76 EKPQSTISGLGLTLYAEVGDIIKVHFKKADKPLSIHPQGIKYSKLSGASVLDHTFPA 135
QY 122 EKEDDKVFPQGSHTYVYVWLKENGPMASDPLCLTYSLVSHVDIVKDLNSGLIGALLVCRE 181
Db 136 EKMDADVAPORETYEWSISEDSGPTDDPPCPTHYIYSHENLIEDFNSGLIGLPLICK 195
QY 182 GSLAKETQ-TLHK-FILLFAVDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVN 239
Db 196 GTLTEGQTKTQKQIVLLFAVEDESKSWSSQSS-----LMTVNGYVN 239
QY 240 RSLPGLIGCHRSKVVYWHVIGMTPTPEVHSIFLGHTFLVRNHRQASLEISPIFLTAQTL 299
Db 240 GTMPDITVCAHDHISWHLMLGSSGPELFIHFNGQVLEQNHKVSALTLLVSASTTANMT 299
```









[illegible]



```

Db 261 TFOENRMAINGVFGNDPELNMCAQKRVAMHLFGMGNEIDVHTAFHGMQLTRGHT 320
QY 284 ASLEISPTFLTAQTLMDLQGLFLFCHISSHQHDMAYVYKVDSCPEPQLRMKNNEA 343
Db 321 DVANIFPAFTVAEVPWEPGTMWLSQVNSHFRDMQALYKVKSCSMAPPV----- 372
QY 344 EDYDDLLTSEMMDVVRPDDNSPSFIQIRSAVKHPKTVWVYIAAEEDWDYAPL----- 398
Db 373 -----DLTGT-----KVRQYFIEAHEIQWYDGMGHDGS 401
QY 399 -----VLAPDRYSKQYLNNGPQRIGRYKVRMAYTDETPKTRTALQHS-----GILGPL 452
Db 402 TGNLREPG--SISDKFFKSSSRIGGYKVRFAQDETFOEKWHEEDHGLGPLV 459
QY 453 LYGEVGDILLIFKQASRPYNYPHGTTDVRPLYSRLPKGVKHLKDFPILPGIFK-- 510
Db 460 IRAEVGDTTQVVFYNRASQPFMSQPHGV-----FYEKDYEGTV--YNDGSSYGLVAKPF 512
QY 511 -----YKWTVTVEDGPTKSPRCITRYYSFVNMERDLASGLIPILLICYKESVDQGNQI 566
Db 513 EKVYRWTPPHAGPTAQDPACLTWYFSAADPIRDTNSGLVGPLVCRAGALGADGKQK 572
QY 567 MSDKRVILFVDENRSLTENIORFLPNPAGVQL-----EDPE-FQASNMHSING 619
Db 573 GVDKEFFLFTVLDEKNSYSN-----ANQAAAMLDFLLSIEDIEGFQDSNRMAING 625
QY 620 YPDSI-QLSVCLHEVAYVYIISGAQTDFLSVFSGYTFKKWYVEDTLTLPSPGETV 678
Db 626 FLFSNLPRLDMCKGDTVAHLLGLGTETDVHGMFQGNVTQVQGMKGAAMLFPHTFMA 685
QY 679 FMSMNPGLWILGCHNSDFRNGMTALLKVSSCDKNTGTYEDSYEDISAYLLSKNNAIE 738
Db 686 IMQPDNLGTFEYCOAGSHREAGRAIYNVSC----- 718
QY 739 PRFSQNPVLRKHQREITRTTLQSDQEBIDYDITISVEMKXEDFDIYDENQSPRSFQ 798
Db 719 -----PGHQ-----ATPRQY 729
QY 799 KTRHYFIAAVERLDYGMSS--PHVLNRAQSGSV-----PQKKVVFQEF 844
Db 730 QAARIYIMAEVENDYCPDRWERHNSQSKSYGIFLNSKDLGLSGRYKKAVFREY 789
QY 845 TDGSGFTQPLRYGELNHLGILGPIYIRAEVEDNIMVTFRQASRPYSFYS-SLISYEEDQR 903
Db 790 TDGTFRIPRPTGPEEHLGILGLKEVGDIITVVFKNARSYPYSVHAHGVLESTTWP 849
QY 904 QGAEPKRFVKNETTYFKVQHMAPTYKDEDFCKAMAYFSDVDLEKDVHSGILGPLLV 963
Db 850 LAEE-----FGEVVTYQWNIPEPSGPGFNDSCVSWIYYGAVDPKIDMYSGLVGLAI 902
QY 964 CHNTLNPAHGRQVTVOEFALPTTIFDETWSYFTENM-ERNCRAPCNITOMEDPTFKENY 1022
Db 903 CQKGLLEPHGSDMDREFFALLFLIDENKSWYLENVAITHGSDQPGSLNLODETFLSN 962
QY 1023 RPHAINGYIMDTPLGLVMAQDRIYLLSMGNSNENIHSIFSGHVTFVRKCEYKMAKY 1082
Db 963 KMHAINGLVANLGLTMOGERVAMVYMLANGQVDLHTIHFABSFYLRNGENYADV 1022
QY 1083 NLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTFLVYNSKCTPLGMASCHIRDFQ 1142
Db 1023 DLEPGTFVEVEMVNASPGTWMHCHVTDVHAGMETLFTVFSR-----TEHLSPLT 1073
QY 1143 ITASQGVQWAPKARLHYSGSINAWSTKEPSWIKVDLLAPMIHGIKTQAGKQFSSL 1202
Db 1074 VITKEKAVPPRDIE---EGNVKMLGMQIPK--NVEMLASVIV-----AI 1115
QY 1203 YISQFIMYSLDGKQWOTVR 1222
Db 1116 SVTLVLLVLAJGGVWVYQHR 1135

```

RESULT 10

```

O75180
ID O75180 PRELIMINARY; PRT; 1104 AA.
AC O75180;
DT 01-NOV-1998 (T-EMBLrel. 08, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE KIAA0698 protein (Fragment).
GN KIAA0698;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98403880; PubMed=9734811;
RA Ishikawa K., Nagase T., Suyama M., Miyajima N., Tanaka A., Kotani H.,
RA Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro."
RL DNA Res. 5:169-176 (1998).
DR EMBL; AB014598; BAA31673.2; -.
DR HSP; P00450; IKNW.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR002355; MultiCu oxidase2.
DR Pfam; PF00394; Cu-oxidase; 3.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
DR PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.
FT NON TER 1
SQ SEQUENCE 1104 AA; 124257 MW; 617C753F766AA152 CRC64;

Query Match 22.2%; Score 1710; DB 4; Length 1104;
Best Local Similarity 32.3%; Pred. No. 4.3e-119;
Matches 397; Conservative 183; Mismatches 432; Indels 216; Gaps 25;

QY 46 YKTLFVEFTVHLFNIAKPRPPMGLGPTQAEVYDVTVTILKNMASHPVSLHVGVS 105
Db 19 YKTIYKEYKDDSYTDEVAQAPAMGLGFLGVLAQEVGDVILLHLKNPATPYTHPHGV 78
QY 106 WKASEGAEYDDQTSOREKEDDKYFPGSGHYTVQVLEKENGPMASDPCLCTYSYL 165
Db 79 EKDSGSLYPDGSGGLKADDSVPFGSHYNTWITPEGHAPTADPACTWIYHSHVDAP 138
QY 166 KDLNLSGILGALLVCREGSL-----AKETQTLHKFILLFAVFDGKSWH-----SETKNS 215
Db 139 RDIATGLIGLITCKRGALDGNPQQQVDVDDHDFLLFSVDENLSWLNENIATYCS 198
QY 216 LMQRDAASARAPKMTVNGYVNRSLPGLIGCHRSVYVHWI GMGTTPEVHSIFLE 275
Db 199 ASVDKEDETFQESNRMAINGFVGNLPELNMCAQKRVAMHLFGMGNEIDVHTAF 258
QY 276 FLVNRHQASLEISPTFLTAQTLMDLQGLFLFCHISSHQHDMAYVYKVDSCPE 335
Db 259 LTRGHTDVANIFPAFTVAEVPWEPGTMWLSQVNSHFRDMQALYKVKSCSMAPP 318
QY 336 RMKNNEAEYDDDLTSEMMDVVRPDDNSPSFIQIRSAVKHPKTVWVYIAAEED 395
Db 319 -----DLTGT-----KVRQYFIEAHEIQWY 339
QY 396 APL-----VLAPDRYSKQYLNNGPQRIGRYKVRMAYTDETPKTRTALQHS 446
Db 340 GPMGHDGSGTKNLEPG--SISDKFFKSSSRIGGYKVRFAQDETFOEKWHEEDH 397
QY 447 --GILGPLLYGEVGDILLIFKQASRPYNYPHGTTDVRPLYSRLPKGVKHLKDF 504
Db 398 HLGLGVVIRAEVGDITQVVFYNRASQPFMSQPHGV-----FYEKDYEGTV-- 450
QY 505 PGEIFK-----YKWTVTVEDGPTKSPRCITRYYSFVNMERDLASGLIPILLI 558
Db 451 PGLVAKPFVKTYRWTPPHAGPTAQDPACLTWYFSAADPIRDTNSGLVGPLVCR 510
QY 559 VDQGNQIMSDKRVILFVDENRSLTENIORFLPNPAGVQL-----EDPE-FQAS 611

```

511	Db	LGADGKQKGVDEKFFLLFTVLNENKSWYN-----ANOAAMLDFFRLSDEIEGFQDS	563
612	Qy	NIMHSINGVYFDSL-OLSVCLHEVAYVYILSICAQTDFLSVFFSGYTFXHKWVYEDTLTL	670
564	Db	NRMHAINGLFNLPRLDCKGDTVAWHLGLGTETDVHGVFMFQGTNVQLQGRKGAAML	623
671	Qy	FPPSGETVPMSENGLWILGCHNSDFPRNGMTALLKVSSCDKNTGDYVDESDVEDISAYL	730
624	Db	FHTFWALMQPDNLGTFRIYCOAGSHRAGRAIYVNSQC-----	664
731	Qy	LGNNAIEPRFSFQNPVPLKRRHOREITRTLQSDQBEIDYDTISVEMKCKEDFIYDEDE	790
665	Db	-----PGHQ-----	668
791	Qy	NOSPRSFQKTRHYFTAAVERLWDYQSSS-PHVLNRAQSGSV-----PQF	836
669	Db	-ATPRQYQAARIYYIMAEVEWMDYCPDRSWEKWEHNSQEKSYGYFISNKGGLGSGRY	727
837	Qy	KKVVOEFTGDGFTQYLGELNEHLLGLGPYIRAEVEONIMWTFRQASRYSFYS-SL	895
728	Db	KKAVFREYTDGTPRIEPRPTGPEHLGILGLIKGEVGDILTWFVKNNASRPYSVHAHGV	787
896	Qy	ISYEEDORQAGPRKNFVKNETKTYFWKVQHHPAPTKEFCKAKAWFSDVDLEKDVHS	955
788	Db	LSGTTVWPLARE-----PGEVVTYQWNIPEKSGPNDACSYYIYSAVDPIKDMTS	840
956	Qy	GLIGPLVCHTNTLNAHQKQVTVQBFALFTTFDETKSWYFTENM-ERNCRAPCNOME	1014
841	Db	GLVGPLAIQCKGILEPHGGRSDMDREFALLFLIFDENKSWYLEENVATHSGSDPGSINLQ	900
1015	Qy	DPTFKENYFHAINGYIMDTLPLGVNAQDORIRWILLSMGSENTHSIHFSHGHVTVRKK	1074
901	Db	DETFLESNKGHALNGLYANLRLGTWYQSERVANYMLAQDQVDLHTIHFHAEFLYIRNG	960
1075	Qy	EYRKALYNLYPGVFTVEMLPSKAGIWRVECLIGELHAGMSTLFLVYSNKRQOTPLGMA	1134
961	Db	ENYRADVDVLFPGTFEWMVWASNPCTWLMHCHVTVDHVAGMETLFTVFSR-----	1011
1135	Qy	SGHIRDFQJATASQYQCMQAPKLARLHVSINAWSTKEPFSWTKVDLLAPMIHGIKTQG	1194
1012	Db	TEHLSPLTWITKETKAVPPRDIE---EGNVQMLGMQIPIK--NVEMLASLV-----	1059
1195	Qy	ARQKFSLSYISOFIMYSLDGKKQWTRY	1222
1060	Db	-----AISVTLILLVVALGVVYVYQHR	1081

RESULT 11	Q9JL97	PRELIMINARY;	PRT; 1084 AA.
ID	Q9JL97		
AC	Q9JL97;		
DT	01-OCT-2000 (TREMBLrel. 15, Created)		
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE	GPI-anchored ceruloplasmin.		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
OX	NCBI_TaxID=10116;		
RN	[1]		
SEQUENCE FROM N.A.			
RP	STRAIN=Sprague Dawley; TISSUE=Brain;		
RC	MEDLINE=20127919; PubMed=10560599;		
RX	Patel B.N., Dunn R.J., David S.;		
RA	"Alternative RNA Splicing Generates a Glycosylphosphatidylinositol-		
RT	anchored Form of Ceruloplasmin in Mammalian Brain.";		
RL	J. Biol. Chem. 275:4305-4310(2000).		
RL	EMBL; AF202115; AAF34175.1; --		
DR	HSSP; P00450; IKCW.		
DR	InterPro; IPR001117; Cu-oxidase.		
DR	InterPro; IPR002355; MultiCu oxidase2.		
DR	Pfam; PF00394; Cu-oxidase; 3.		

```

Db 849 RYIIQIPERSAGTSDPCIPWALYSTVDRVKDYISGLIGLIVCRKSYKVNPK--- 905
QY 976 QYVQEFALFTIFDTKSWYFTENMERNCRAPCNIOMEDPTFKENYRFAHNGYIMDTL 1035
Db 906 --KWEYSLLFLVFDENESWYLDNINTYSDHPEKVNKNBEFTESNKNHAIKNGKMGFNL 963
QY 1036 PGLVMAQDQIRWYLLSGNSNENIHSIHFSGHVFTVRKKEYSKALYNLYGVETVEML 1095
Db 964 QGLTHVGVGVNMYWAMGNEIDLTHVHFHGHSPQYKRGHSHSDVDFLPFGTYQTLMP 1023
QY 1096 PSKAGIWRVECLIGELHAGMSTLFLVYSNKKQT 1129
Db 1024 PQTGPGWLLHCHVTDHAGMVTYTVLPNQASS 1057

RESULT 12
Q9XT27
ID Q9XT27 PRELIMINARY; PRT; 1048 AA.
AC Q9XT27;
DT 01-NOV-1999 (TREMELrel. 12, Created)
DT 01-NOV-1999 (TREMELrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)
DE Ceruloplasmin.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=99384006; PubMed=10452945;
RA Lockhart P.J., Mercer J.F.B.;
RT "Cloning and expression analysis of the sheep ceruloplasmin cDNA.";
RL Gene 236:251-257(1999).
DR EMBL; AF134814; AAD41477.1; -.
DR HSSP; P00450; 1KCW.
DR InterPro; IPR001117; Cu-oxidase.
DR Pfam; PF00394; Cu-oxidase; 3.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
DR PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.
SQ SEQUENCE 1048 AA; 119125 MW; 925F16D7B0549CBB CRC64;

Query Match 21.0%; Score 1618; DB 6; Length 1048;
Best Local Similarity 32.2%; Pred. No. 3.2e-112;
Matches 378; Conservative 177; Mismatches 417; Indels 202; Gaps 25;

QY 3 RRYILGAVLSWYMSDLGE---LPVDARPPRPVKSPFPNTSVYKKTFLVEFTVILF 59
Db 22 KHYTIGIETAWNY-ASDHAEEKLISVDTHSNYIYLONGPNRIGSVYKKAVALQYTDNF 80
QY 60 NIAKPRPPMGLIGPTIOAEVDTVITLKNASHPSVLSHAGVSYKASGAEVDDOTS 119
Db 81 RTVIEKFWLGFGLFIKAETGDKVTVHLKXNPASPYTFHAGLTYHEGALYPDNTT 140
QY 120 QREKEDKVPFGSSHTYVWQVLKNGPMASDPLCTLYLSHVDLVKDLNSGLICALVVC 179
Db 141 DLQAKADKVPQFQCLYILHANPEQPGBEDSNCTVTHSHIDAPKDIAASGLIGLTHC 200
QY 180 RSGSLAKEKTQTLHK-FILLEAVFDEGKSWH-----SETXNSLMQDDAASARAWP 229
Db 201 KQSLDEEKEKIDKEFVVMFVVDENUSWYLEENIKYCSPEKVEQDNEDFQESN--- 257
QY 230 KMHTVNGYVNRSLPGLIGCHRSVYVHWVIGMTTPEVHSIFLEGHTFLVRNHRQASLEIS 289
Db 258 RMYSVNGYAFGLSLPGLSMCAEDRVKWLFGMGNEIDVHAFFHGQVLSKNVRVDTINLF 317
QY 290 PITFLTAQTLMDLQGLFLFCHISSHQHDGMAYVKVDSCEPPEQLRMKNNEEABDYDD 349
Db 318 PATLFDAFVVAQNPQGMWLSQNLNHLKAGLQAFWVQDCKSS----- 361

```

```

QY 350 LTDSEMDVVRFDSDNSPSFIQIRSVAKGPKTWVHYIAAREEDWDYAPL-----VL 400
Db 362 -----SEDN-----IHGKNVRH-----YIIAAEEVIMWYAFSGIDAFKEMLR 399
QY 401 APDRSYKSYQYLNNGSPQIRKRYKVRFMAYTDTF---KTRERAIQHSGLIGLPLLYGEV 457
Db 400 APGSAS--EAFPEQGPTRIGGSYKXKLVRREYTDASFNSQKGRGPEEHLGILGPVIAEV 457
QY 458 GDTLLIIPKQASRPNIYPHGI-----TDVRLPSRRLPKGVKHLKOPPIIPGIF 509
Db 458 GDTIRVTFNKAAHPUSIEPIGVVRVDKNNEGTYYSPTGSGPPSG-SH-----VAPKGT 511
QY 510 KYKWTIVTVDGPTKSDPCLTRYISFVNMERDLAGLIGLPLLYCVKESVDQGNQIMSD 569
Db 512 TYEMTVPEVGYTKDPVCLAKMYIS--GSTKDITGLIGPMKI CNGSLLANGRLKXND 569
QY 570 KRNVIILSPVDENRSWYLTENIQRFLPNPAGVOLDEPBPQASINIMHSINGYVDSLQ-LS 628
Db 570 KEFYLPFTVFDENESLLDDNIMFTTAPDQVDKENEDFQESNKMHSNMNGFYGNQGLS 629
QY 629 VCLHEVAVWYILSIGAQTDFLSVFFSGYTFKHKWVYEDTLTPFSGETVFMMSMENPGLW 688
Db 630 MCQGSVMWYILFSAGNEVDIHGIYFSGNTYILRGERRDTANLPQTSLSLFPQDTAGTF 689
QY 689 ILGCHNSDFRNRGMTALLKVSXCDKNTGDIYEDSYEDISAYLLSKNNAIEPRFSQNPV 748
Db 690 DVECLATDHTVTGGMQKYTVSQGQRS-----EDLYLYL----- 723
QY 749 LKXHQREIRTTLOSQDEIDYDDTISVEMKEDFDIYDEDNQSPRSFQKTRHYFTAA 808
Db 724 -----CERTYYTAA 732
QY 809 VERLWDYGMS-----SSPHVLRNRAOAGSV-----POPKKVVFOEFTDGSFTOPLYR 855
Db 733 VEVEWDYSPSRKWEKELHLQSNLSNAFLDKKEFYIGSKYKVVYRQFTDSTFQVYVER 792
QY 856 GEINELHGLIGPYIRAEVDNIMVTRNQAASPYFYSSLIYSIEDQOQASPRKNFKVP 915
Db 793 KGEEHLGILGPQLHADVGDKVNIIPKKNATPRSYTHAHGVKTESST---VTP---TAP 845
QY 916 NETKTYFMKVOHMAPTKDEEDCKAWAYPSDDVLEKDVHSGLIGLPLVCHTWTL---NPA 972
Db 846 GEIRTTINKLIPERSGAGMDSPCIWVYISTVDRVKDLFSGLIGLIVCRKHYLVSNP- 904
QY 973 HGRQVTVQEFALFTIFDTKSWYFTENMERNCRAPCNIOMEDPTFKENYRFAHNGYIM 1032
Db 905 ----IKKLEFSLFLVFDENESWYLDNINTYSDHPEKVDKANEEFMESNKMHAINGRMP 960
QY 1033 DTLPLGLVMAQDQIRWYLLSGNSNENIHSIHFSGHVFTVRKKEYSKALYNLYGVETV 1092
Db 961 GNLQGLTM-----HVGNEVDLSHVHFGHSPQYHGRGIYTSVDVDFLPFGTYQTL 1009
QY 1093 EMLPSKAGIWRVECLIGELHAGMSTLFLVYSNK 1126
Db 1010 EMTPTPGIWLHCHVTDHAGMVTYTVLPNE 1043

RESULT 13
Q8BQ43
ID Q8BQ43 PRELIMINARY; PRT; 355 AA.
AC Q8BQ43;
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE Coagulation factor VII (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Boreal root ganglion;
RX MEDLINE=22354683; PubMed=12466851;

```





